Implicit Runge-Kutta methods for general linear index 2 differential-algebraic equations with variable coefficients

Michael Hanke and Ebroul Izquierdo Macana
Fachbereich Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
D–10099 Berlin
Germany

Abstract

We investigate the approximation of linear index 2 differential-algebraic equations by implicit Runge-Kutta methods. Implicit Runge-Kutta methods cannot be feasible or become unstable when applied to linear differential-algebraic equations with variable nullspace. In order to overcome these difficulties we propose a new approach for this case. The feasibility of the weak instability and convergence are proved. Order results are given in terms of the Butcher identities.

Key words: ordinary differential equations, differential-algebraic equations, initial value problems, implicit Runge-Kutta methods.
Mathematics Subject Classification: 65L10

Running head: Runge-Kutta methods for index 2 dae's

1. Introduction

In this paper we consider the numerical approximation of initial value problems for linear differential-algebraic equations (DAEs) of the form

$$A(t)x'(t) + B(t)x(t) = q(t). \tag{1.1a}$$

We assume that the coefficients A, B in (1.1a) are continuous matrix functions $A,\ B : [t_0, T] \longrightarrow L(I\!R^m)$ and the matrix $A(t)$ has a smooth nullspace $N(t)$ for $t \in [t_0, T]$, i. e. there are continuously differentiable projection functions Q, P : $[t_0, T] \longrightarrow L(I\!R^m)$ so that $Q(t)$ is a projection onto the nullspace $N(t)$ and $P(t) \equiv I - Q(t)$. Note that a smooth nullspace $N(t)$ has always constant dimension on $[t_0, T]$ and $A(t)$ has constant rank. Such equations arise in a variety of applications, e. g. electrical network theory, dynamical systems subject to constraints.

In the present paper, we are intersted in index 2 equations (more precisely, DAEs being tractable with index 2). Besides the BDF, implicit Runge-Kutta methods are common for approximating initial value problems in DAEs. Since higher index DAEs are known to be ill-posed, some kind of instability should be expected. Nevertheless, some IRK methods are reported to be appropriate for some special types of higher index equations. Especially so-called Hessenberg systems are well investigated [1], [2], [4], [7]. We will generalize these results to the general linear DAE (1.1a). In order to overcome the inherent difficulties when the nullspace of A varies we introduce a new approach for IRK methods using the projection Q. Doing so we can show that these methods share all properties that IRK methods have when being applied to Hessenberg systems even for general DAEs (1.1a).

Following the idea of März [10] we use the matrix functions

$$\begin{aligned} G_1 &:= A + BQ, \\ A_1 &:= G_1 - AP'Q, \\ G_2 &:= A_1 + BPQ_1, \end{aligned}$$

as well as the subspaces

$$\begin{aligned} N(t) &:= \ker(A(t)), \\ S(t) &:= \{x \in I\!R^m \ : \ B(t)x \in \operatorname{im}(A(t))\}, \\ N_1(t) &:= \ker(A_1(t)), \\ S_1(t) &:= \{x \in I\!R^m \ : \ B(t)P(t)x \in \operatorname{im}(A_1(t))\}. \end{aligned}$$

Thereby, $t \in [t_0, T]$ and $Q_1(t)$ denote a projection onto $N_1(t)$, further $P_1(t) := I - Q_1(t)$. The subspaces N, S, N_1 and S_1 are called canonical subspaces of the DAE (1.1a).

Definition 1.1 *(März [11]) The DAE (1.1a) is called transferable if $G_1(t)$ is nonsingular for every $t \in [t_0, T]$. The DAE (1.1a) is called tractable with index 2 if $G_1(t)$ is singular, but $G_2(t)$ is nonsingular for every $t \in [t_0, T]$.*

Note that $G_1(t)$ is nonsingular iff $A_1(t)$ is so. The transferable DAEs are well understood. For transferable DAEs suitably modified numerical ODE methods work well. We refer to [2], [4] for this case. Nontransferable DAEs are essentially more complex. Tractability with index 2 characterizes an important class of nontransferable DAEs. The maybe most important qualitative difference between transferable and nontransferable DAEs is that initial and boundary value problems for nontransferable DAEs are ill-posed in the sense of Tikhonov whereas the first class of problems remains well-posed in the Hadamard sense. This fact has to be taken into account for the numerical treatment.

The formulation of appropriate initial conditions for index 2 tractable DAEs depends on the DAE itself. The index 2 tractable DAE (1.1) subject to the condition

$$P(t_0)P_1(t_0)x(t_0) = b \tag{1.1b}$$

represents an initial value problem (IVP) with consistent initial values.

According to the originally conceived method for the numerical solution of ordinary differential equations (Kutta, 1901), an IRK method can be realized for the DAE (1.1a) in the following way:
From an approximation x_{n-1} of the solution of the IVP (1.1) at t_{n-1} these one-step methods construct an approximation x_n at $t_n = t_{n-1} + h$ via the formula

$$x_n = x_{n-1} + h\sum_{j=1}^{s} b_j X_j', \tag{1.2a}$$

where X_j' is defined by

$$A(t_{nj})X_j' + B(t_{nj})X_j = q(t_{nj}), \ \ j = 1, 2, \ldots, s \tag{1.2b}$$

with $t_{nj} := t_{n-1} + c_j h$ and internal stages X_j given by

$$X_j = x_{n-1} + h\sum_{l=1}^{s} a_{jl}X_l', \ \ j = 1, 2, \ldots, s. \tag{1.2c}$$

Here a_{jl}, b_j, c_j are the coefficients which determine the method, and s is the number of stages. Usually, the matrix $\mathcal{A} = (a_{jl})_{j,l=1}^{s}$ and the vectors $b = (b_1, b_2, \ldots, b_s)^T$, $c = (c_1, c_2, \ldots, c_s)^T$ are combined in the so-called Butcher Diagram [3] or Runge-Kutta scheme

$$\begin{array}{c|c} c & \mathcal{A} \\ \hline & b^T \end{array} .$$

The paper is organized as follows. In section 2 we give conditions which ensure the feasibility of IRK methods. These conditions depend not only on the method but essentially on the behaviour of the nullspace $N(t)$. This is similar to what is known for the BDF [11]. In section 3 we consider the case of a constant nullspace $N(t) \equiv N$. We show the weak instability and give order results in terms of the Butcher identities. A well-known simple example shows that similar results are not true if the nullspace is variable. Therefore, we introduce a new approximation for this case which makes use of the projection Q. In section 4 we show that these new approximations share their properties with the original IRK methods in the constant nullspace case. Section 5 contains some numerical examples.

Assumption: *In the following, $Q_1(1)$ denotes the canonical projection onto $N_1(t)$ along $S_1(t)$. Moreover, we assume that all matrix functions and projections are sufficiently smooth.*

2. Existence and uniqueness of Runge-Kutta solution

In a first step we study the existence and uniqueness of a Runge-Kutta solution, i. e. we try to give an answer to the question: When does the system (1.2b) have a unique solution X_j (or X_j') for $j = 1, 2, \ldots, s$?

Supposed that the Runge-Kutta matrix $\mathcal{A}$ is nonsingular, the IRK method (1.2) is equivalent with

$$x_n = \rho x_{n-1} + \sum_{j=1}^{s}\sum_{l=1}^{s} b_j \hat{a}_{jl} X_l, \tag{2.1a}$$

$$A(t_{nj}) \sum_{l=1}^{s} \hat{a}_{jl}(X_l - x_{n-1}) + hB(t_{nj})X_j = hq(t_{nj}), \quad j = 1, 2, \ldots, s, \tag{2.1b}$$

where $\mathcal{A}^{-1} = (\hat{a}_{jl})_{j,l=1}^{s}$ and $\rho := 1 - \sum_{i=1}^{s}\sum_{j=1}^{s} b_i \hat{a}_{ij}$ ([8]).
We consider now the IRK method defined by (2.1). Let us introduce some notations.

$$\begin{aligned} D_A &= \operatorname{diag}(A(t_{n1}), \ldots, A(t_{ns})), \;\; D_B = \operatorname{diag}(B(t_{n1}), \ldots, A(t_{ns})), \\ D_q &= \operatorname{diag}(q(t_{n1}), \ldots, q(t_{ns})), \;\; \mathbb{1}_s = (1, \ldots, 1)^T \in I\!R^s, \\ \bar{X} &= (X_1^T, X_2^T, \ldots, X_s^T)^T \in I\!R^{ms}. \end{aligned}$$

Now the system (2.1b) can be written in the form

$$[D_A(\mathcal{A}^{-1} \otimes I_m) + hD_B]\bar{X} = hD_q \mathbb{1}_s + D_A(\mathcal{A}^{-1}\mathbb{1}_s \otimes x_{n-1}). \tag{2.2}$$

The IRK method (2.1) is feasible iff the matrix $\mathcal{M} := D_A(\mathcal{A}^{-1} \otimes I_m) + hD_B$ is nonsingular.
Unfortunately, simple examples show that $\mathcal{M}$ is not always so.

Example 2.1 Consider

$$A(t) = \begin{pmatrix} 0 & 0 & 0 \\ t & e^{-t} & 0 \\ 0 & 0 & t \end{pmatrix}, \quad B(t) = \begin{pmatrix} t & e^{-t} & 0 \\ 0 & 0 & 3 \\ 0 & 0 & t+1 \end{pmatrix}.$$

The related DAE is transferable with index 2 for $t \in (0, \infty)$ and

$$Q = \begin{pmatrix} 0 & -e^{-t}/t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is a projector onto $N(t) = \ker(A(t))$. For a generic 1-stage IRK method this yields

$$\mathcal{M} = A(t)(a^{-1} \otimes I_3) + hB(t) = \begin{pmatrix} ht & he^{-t} & 0 \\ a^{-1}t & a^{-1}e^{-t} & 3h \\ 0 & 0 & a^{-1}t + ht + h \end{pmatrix}.$$

Obviously, this matrix is singular for all $t, h \in (0, \infty)$ and all $a \in \Re \setminus \{0\}$. $\square$

The following theorem provides a condition for the existence of $\mathcal{M}^{-1}$.
Throughout this paper we denote $D_N = \operatorname{diag}(N(t_{n1}), \ldots, N(t_{ns}))$ for any matrix function $N(t)$.

Theorem 2.1 *Suppose that the IVP (1.1) is tractable with index 2 and, moreover, that $\mathcal{A}$ and $I_{ms} - \mathcal{E}D_{QQ_1P'}$ are nonsingular for all $t \in [t_0, T]$, where $Q_1(t)$ denotes the canonical projection onto $N_1(t)$ along $S_1(t)$. Then the system (2.2) is uniquely solvable for sufficiently small h.*
Here $\mathcal{E} := I_{ms} + [(\mathcal{C} \odot \mathcal{A}^{-1}) \otimes I_m]$, $\mathcal{C} = (c_{jl})_{j,l=1}^{s} \in L(I\!R^s)$, $c_{jl} = c_j - c_l$ and $\mathcal{C} \odot \mathcal{A}^{-1} = (c_{jl}\hat{a}_{jl})_{j,l=1}^{s} \in L(I\!R^s)$.

Proof Let $D := I_{ms} - \mathcal{E}D_{QQ_1P'}$. (2.2) is equivalent with

$$\begin{aligned}&[D_{P_1P}(\mathcal{A}^{-1} \otimes I_m) + hD_{P_1PP'Q} + hD_{P_1Q} + hD_{Q_1} + hD_{G_2^{-1}BPP_1}]\bar{X} = \\ &hD_{G_2^{-1}q}1\!\!1_s + D_{G_2^{-1}A}(\mathcal{A}^{-1} \otimes I_m)(1\!\!1_s \otimes x_{n-1}).\end{aligned}$$

Multiplying this equation by D_{PP_1}, $\frac{1}{h}D_{QP_1}$ and $\frac{1}{h}D_{Q_1}$ and using the Taylor expansion we obtain the system

$$\begin{aligned}((\mathcal{A}^{-1} \otimes I_m) &+ hD_{PP_1G_2^{-1}B} + R_{PP_1})D_{PP_1}\bar{X} + R_{PP_1}D_{PQ_1}\bar{X} + (hD_{PP_1P'} + R_{PP_1})D_Q\bar{X} = \\ &hD_{PP_1G_2^{-1}q}1\!\!1_s + D_{PP_1}(\mathcal{A}^{-1} \otimes I_m)(1\!\!1_s \otimes x_{n-1}), &(2.3a)\\ D_{Q_1}\bar{X} &= D_{Q_1G_2^{-1}q}1\!\!1_s, &(2.3b)\\ D_{QP_1G_2^{-1}BPP_1}\bar{X} &- \frac{1}{h}(\mathcal{A}^{-1} \otimes I_m)D_{QQ_1}\bar{X} + (I_{ms} - D_{QQ_1P'})D_Q\bar{X} \\ &= D_{QP_1G_2^{-1}q}1\!\!1_s - \frac{1}{h}D_{QQ_1}(\mathcal{A}^{-1} \otimes I_m)(1\!\!1_s \otimes x_{n-1}) + \frac{1}{h}R_{QQ_1}\bar{X}, &(2.3c)\end{aligned}$$

where $R_{PP_1} = h[((\mathcal{C} \odot \mathcal{A}^{-1}) \otimes I_m)D_{(PP_1)'} + O(h)]$, and

$$\begin{aligned}\frac{1}{h}R_{QQ_1}\bar{X} &= ((\mathcal{C} \odot \mathcal{A}^{-1}) \otimes I_m)D_{(QQ_1)'} + O(h))\bar{X} \\ &= ((\mathcal{C} \odot \mathcal{A}^{-1}) \otimes I_m)D_{(Q'Q_1+QQ_1')}((D_{PP_1} + D_{PQ_1} + D_Q + O(h))\bar{X}.\end{aligned}$$

Inserting (2.3a) and (2.3b) into the latter equation and using the relations $Q'Q_1PP_1 = 0$ and $Q_1Q = 0$ we obtain

$$\frac{1}{h}R_{QQ_1}\bar{X} = \mathcal{V} + ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'Q}\bar{X} + O(h\|\bar{X}\|), \tag{2.4}$$

$$\begin{aligned}\text{where}\quad \mathcal{V} \;=\; & ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'}[h(\mathcal{A}\otimes I_m)D_{PP_1G_2^{-1}q}1\!\!1_s + \\ & (\mathcal{A}\otimes I_m)D_{PP_1}(\mathcal{A}^{-1}\otimes I_m)(1\!\!1_s\otimes x_{n-1})] + \\ & ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{(QQ_1)'PQ_1G_2^{-1}q}1\!\!1_s\ .\end{aligned}$$

We insert (2.4) into (2.3c). The equation obtained in this way together with (2.3a-b) can be written as

$$\mathcal{F}\begin{bmatrix} D_{PP_1}\bar{X} \\ D_{Q_1}\bar{X} \\ D_Q\bar{X}\end{bmatrix} = \mathcal{J} + \begin{bmatrix} O(h) & O(h) & -h\mathcal{S}+O(h^2) \\ 0 & 0 & 0 \\ O(h) & O(h) & O(h)\end{bmatrix}\begin{bmatrix} D_{PP_1}\bar{X} \\ D_{PQ_1}\bar{X} \\ D_Q\bar{X}\end{bmatrix}, \tag{2.5}$$

where

$$\mathcal{F} \;=\; \begin{bmatrix} I_ms & 0 & 0 \\ 0 & I_{ms} & 0 \\ D_{QP_1G_2^{-1}B} & -\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_Q & \mathcal{D}\end{bmatrix},$$

$$\mathcal{J} \;=\; \begin{pmatrix} h(\mathcal{A}\otimes I_m)D_{PP_1G_2^{-1}q}1\!\!1_s + (\mathcal{A}\otimes I_m)D_{PP_1}(\mathcal{A}^{-1}\otimes I_m)(1\!\!1_s\otimes x_{n-1}) \\ D_{Q_1G_2^{-1}q}1\!\!1_s \\ D_{QP_1G_2^{-1}q}1\!\!1_s - \frac{1}{h}D_{QQ_1}(\mathcal{A}^{-1}\otimes I_m)(1\!\!1_s\otimes x_{n-1}) + \mathcal{V}\end{pmatrix},$$

and $\mathcal{S} = [(\mathcal{A}\otimes I_m) + (\mathcal{A}(\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m]D_{(PP_1)'Q}$.
Note that in (2.5) the identity $QQ_1'Q = QQ_1'QQ = QQ_1P'Q$ had be used.

The matrix $\mathcal{F}$ is nonsingular iff $\mathcal{D}$ is so and in this case we can compute $\mathcal{F}^{-1}$:

$$\mathcal{F}^{-1} = \begin{bmatrix} I_{ms} & 0 & 0 \\ 0 & I_{ms} & 0 \\ -\mathcal{D}^{-1}D_{QP_1G_2^{-1}B} & \frac{1}{h}\mathcal{D}^{-1}(\mathcal{A}^{-1}\otimes I_m)D_Q & \mathcal{D}^{-1}\end{bmatrix}.$$

Hence, we have from (2.5)

$$\begin{bmatrix} D_{PP_1}\bar{X} \\ D_{PQ_1}\bar{X} \\ D_Q\bar{X}\end{bmatrix} = \begin{bmatrix} I_{ms} & 0 & 0 \\ 0 & D_P & 0 \\ 0 & 0 & I_{ms}\end{bmatrix}\mathcal{F}^{-1}\mathcal{J} + \begin{bmatrix} O(h) & O(h) & O(h) \\ 0 & 0 & 0 \\ O(h) & O(h) & O(h)\end{bmatrix}\begin{bmatrix} D_{PP_1}\bar{X} \\ D_{PQ_1}\bar{X} \\ D_Q\bar{X}\end{bmatrix}. \tag{2.6}$$

Using Banach's lemma we can solve the system (2.6) for sufficiently small h. Now we use the identity $\bar{X} = D_{PP_1}\bar{X} + D_{PQ_1}\bar{X} + D_Q\bar{X}$ to construct the unique solution of (2.2). $\square$

Remarks

1. For 1-stage IRK methods it holds that $I_m - \mathcal{E}D_{QQ_1P'} = I_m - QQ_1P'$, because in this case
$\mathcal{C} = 0$. In order to verify the solvability of the system (2.2) it is sufficient to check the nonsingularity of the matrix $I_m - QQ_1P'$. In example 2.1
$$I_3 - QQ_1P' = \begin{pmatrix} 1 & e^{-t}/t & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
is always singular.

2. In [11] it is proved that BDF methods for the numerical solution of (1.1) are feasible if the matrix $I_m - QQ_1P'$ is nonsingular and h is sufficiently small. For IRK methods with s-stages, $s \geq 2$, the feasibility of the method depends on the method and the IVP.

3. If the nullspace $N(t)$ of $A(t)$ is constant, then, for h sufficiently small, the IRK method is always feasible, because in this case $P' = 0$.

Example 2.2 [2], [8], [11] Let
$$A(t) = \begin{pmatrix} 0 & 0 \\ 1 & \eta t \end{pmatrix}, \qquad B(t) = \begin{pmatrix} 1 & \eta t \\ 0 & 1+\eta \end{pmatrix}$$
with $\eta \in \Re$. The related DAE is transferable with index 2
$$\begin{aligned} Q(t) &= \begin{pmatrix} 0 & -\eta t \\ 0 & 1 \end{pmatrix}, \quad A_1(t) = \begin{pmatrix} 0 & 0 \\ 1 & 1+\eta t \end{pmatrix} \quad \text{and} \\ Q_1(t) &= \begin{pmatrix} 1+\eta t & \eta t(1+\eta t) \\ -1 & -\eta t \end{pmatrix}. \end{aligned}$$
It is known that the backward Euler method is not feasible if $\eta = 1$. If we apply the 2-stage Lobatto IIIC-formula [3] to this DAE instead, we have:
$$\mathcal{A}^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \qquad \mathcal{C} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$
$$\mathcal{E} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad I_{ms} - \mathcal{E}D_{QQ_1P'} = \begin{pmatrix} 1 & -\eta^2 t_{n1} & 0 & \eta^2 t_{n2} \\ 0 & 1+\eta & 0 & -\eta \\ 0 & \eta^2 t_{n1} & 1 & -\eta^2 t_{n2} \\ 0 & -\eta & 0 & 1+\eta \end{pmatrix}.$$
This last matrix is singular iff $\eta = -\frac{1}{2}$. Moreover,
$$\mathcal{M} = \begin{pmatrix} h & h\eta t_{n1} & 0 & 0 \\ 1 & \eta t_{n1} + h(1+\eta) & 1 & \eta t_{n1} \\ 0 & 0 & h & h\eta t_{n2} \\ -1 & -\eta t_{n2} & 1 & \eta t_{n2} + h(1+\eta) \end{pmatrix},$$
and $\mathcal{M}$ is singular iff $\eta = -\frac{1}{2}$. $\square$

Now we will study the stability and convergence of the IRK method (1.2). To that end we consider two cases: The constant nullspace and the nonconstant nullspace.

3. Stability and convergence for the constant nullspace case

IRK methods become unstable when they are applied to DAEs with index 2. This fact is well-known in the simple case of linear DAEs with constant coefficients and constant stepsize. We refer to [2], [8] for this case. For linear DAEs with nonconstant coefficients but constant nullspace $N(t)$, some terms of the system (2.3) get lost, because $P' = 0$, and solving this system with respect to $D_{PP_1}\bar{X}$, $D_{PQ_1}\bar{X}$ and $D_Q\bar{X}$ we obtain

$$\begin{aligned} D_{PP_1}\bar{X} &= h((\mathcal{A}\otimes I_m)+O(h))D_{PP_1G_2^{-1}q}1\!\!1_s - h\Big[(\mathcal{A}(\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{PP_1'} + \\ &\quad O(h)\Big]D_{PQ_1G_2^{-1}q}1\!\!1_s + (D_{PP_1}+O(h))D_P(1\!\!1_s\otimes x_{n-1}), \qquad (3.1a) \\ D_{PQ_1}\bar{X} &= D_{PQ_1G_2^{-1}q}1\!\!1_s\,, \qquad (3.1b) \end{aligned}$$

$$\begin{aligned} D_Q\bar{X} &= \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QQ_1} + D_{QP_1} + ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'PQ_1} + \\ &\quad O(h)\Big]D_{G_2^{-1}q}1\!\!1_s + \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QP_1P} - D_{QP_1G_2^{-1}BPP_1} + \\ &\quad ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'}(D_{PP_1}-I_{ms}) + O(h)\Big]D_P(1\!\!1_s\otimes x_{n-1}). \qquad (3.1c) \end{aligned}$$

Clearly, the nontrivial term $1/h(\mathcal{A}^{-1}\otimes I_m)D_{QP_1P}$ in (3.1c) grows unboundedly if h tends to zero, i. e. the method remains unstable. Nevertheless we can prove convergence. In order to make this fact more transparent we show in the following that in the special case of a constant nullspace the instability is weak and that it refers only to the Q-component. For that purpose we consider in addition to the vectors x_n obtained from the formulas (2.1) the $\tilde{x}_n$ resulting from the perturbed systems

$$\tilde{x}_n = \rho\tilde{x}_{n-1} + \sum_{j=1}^{s}\sum_{l=1}^{s} b_j\hat{a}_{jl}\tilde{X}_l \quad + \quad h\delta_{n,s+1}, \qquad (3.2a)$$

$$\frac{1}{h}A(t_{nj})\sum_{l=1}^{s}\hat{a}_{jl}(\tilde{X}_l-\tilde{x}_{n-1}) + B(t_{nj})\tilde{X}_j \quad = \quad q(t_{nj})+\delta_{nj}, \quad j=1,2,\ldots,s. \qquad (3.2b)$$

Assume $\|\delta_{ij}\| \le \delta$ for all $i=0,1,\ldots,n$, $j=1,2,\ldots,s+1$.
Analogously to (2.2) we can write the system (3.2b) in the form

$$\mathcal{M}\bar{\tilde{X}} = hD_{(q+\delta_n)}1\!\!1_s + D_A(\mathcal{A}^{-1}1\!\!1_s\otimes x_{n-1}),$$

where $\bar{\tilde{X}} = (\tilde{X}_1^T, \tilde{X}_2^T, \ldots, \tilde{X}_s^T)^T$.

According to Theorem 2.1 $\mathcal{M}$ is nonsingular for sufficiently small h, i. e. the system (3.2b) is uniquely solvable.

Let us introduce some new notations:
$\triangle x_k := \tilde{x}_k - x_k$, $\triangle Px_k := P(\tilde{x}_k - x_k)$, $\triangle PP_1x_k := PP_1(t_k)(\tilde{x}_k - x_k)$,
$\triangle PQ_1x_k := PQ_1(t_k)(\tilde{x}_k - x_k)$, $\triangle Qx_k := Q(\tilde{x}_k - x_k)$, $\triangle PP_1X_j := PP_1(t_{nj})(\tilde{X}_j - X_j)$,
$\triangle PQ_1X_j := PQ_1(t_{nj})(\tilde{X}_j - X_j)$, $\triangle QX_j := Q(\tilde{X}_j - X_j) \in I\!R^m$ and $\triangle \bar{X} := \tilde{\bar{X}} - \bar{X} \in I\!R^{ms}$
for $k = n-1, n$, $j = 1, 2, \ldots, s$.

For the difference $\triangle \bar{X}$ we obtain (see (3.1))

$$\begin{aligned}
D_{PP_1}\triangle\bar{X} &= h((\mathcal{A}\otimes I_m) + O(h))D_{PP_1G_2^{-1}\delta_n}1\!\!1_s - h\Big[(\mathcal{A}(\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{PP_1'} + \\
&\quad O(h)\Big]D_{PQ_1G_2^{-1}\delta_n}1\!\!1_s + (D_{PP_1} + O(h))D_P(1\!\!1_s\otimes\triangle x_{n-1}), \qquad (3.3a)\\
D_{PQ_1}\triangle\bar{X} &= D_{PQ_1G_2^{-1}\delta_n}1\!\!1_s\,, \qquad (3.3b)\\
D_Q\triangle\bar{X} &= \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QQ_1} + D_{QP_1} + ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'PQ_1} + \\
&\quad O(h)\Big]D_{G_2^{-1}\delta_n}1\!\!1_s + \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QP_1P} - D_{QP_1G_2^{-1}BPP_1} + \\
&\quad ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'}(D_{PP_1} - I_{ms}) + O(h)\Big]D_P(1\!\!1_s\otimes\triangle x_{n-1}). \qquad (3.3c)
\end{aligned}$$

Using the Taylor expansion about t_{n-1}, and equating like powers of h, it is easy to see that the differences $\triangle X_j$, $j = 1, 2, \ldots, s$, satisfy:

$$\begin{aligned}
\|\triangle PP_1X_j\| &\le \|\triangle PP_1x_{n-1}\| + hC(\|\triangle Px_{n-1}\| + \delta), \qquad (3.4a)\\
\|\triangle PQ_1X_j\| &\le C\delta, \qquad (3.4b)\\
\|\triangle QX_j\| &\le \frac{C}{h}(\|\triangle PQ_1x_{n-1}\| + h\|\triangle Px_{n-1}\| + \delta), \qquad (3.4c)
\end{aligned}$$

with a constant C.

These estimates reflect the influence of pertubations on the solution of (2.1b). In order to prove the main results of this section we need the following lemma. For the proof we refer to [5], [8].

Lemma 3.1 *Suppose that the two sequences of positive numbers (y_n), (z_n) satisfy*

$$\begin{aligned}
y_n &\le y_{n-1} + hC(y_{n-1} + z_{n-1} + D_1),\\
z_n &\le |\,\rho\,|z_{n-1} + hC\Big(y_{n-1} + z_{n-1} + \frac{D_2}{h}\Big).
\end{aligned}$$

Then we have for $hn \le C_0$:

$$\begin{aligned}
y_n &\le C_1(y_0 + hy_0 + hz_0 + D_1 + D_2), \quad \textit{if } |\,\rho\,| \le 1,\\
z_n &\le C_1(hy_0 + (|\,\rho\,|^n + h)z_0 + hD_1 + D_2), \quad \textit{if } |\,\rho\,| < 1,\\
z_n &\le C_1(hy_0 + (1+h)z_0 + hD_1 + \frac{D_2}{h}), \quad \textit{if } |\,\rho\,| = 1.
\end{aligned}$$

C, C_0, C_1, D_1, D_2 are positive constants.

Theorem 3.1 *Let the assumptions of Theorem 2.1 be satisfied and, furthermore, $P' = 0$. If the coefficients of the IRK method (2.1) satisfy the relation $|\rho| = |b^T\mathcal{A}^{-1}1\!\!1_s - 1| < 1$, then the IRK method (2.1) applied to the DAE (1.1) is weakly unstable. This weak instability refers to the Q-component of the numerical solution. The P-component is stable. More precisely, it holds that*

$$\|\triangle Px_n\| \leq C\delta, \|\triangle Qx_n\| \leq \frac{C}{h}\delta.$$

Proof Subtracting (2.1a) from (3.2a), we obtain

$$\triangle x_n = \rho\, \triangle x_{n-1} + (b^T\mathcal{A}^{-1} \otimes I_m)\triangle\bar{X} + h\delta_{n,s+1}. \tag{3.5}$$

Multiplying this equation by $PP_1P(t_n) = PP_1(t_n)$, $PQ_1P(t_n) = PQ_1(t_n)$ and Q, and using the Taylor expansion at t_{n-1} on the right-hand side of the equations obtained in this way we get

$$\begin{aligned}
\triangle PP_1x_n &= \rho\triangle PP_1x_{n-1} + h\rho(PP_1'(t_{n-1}) + O(h))\triangle Px_{n-1} + \\
&\quad (b^T\mathcal{A}^{-1} \otimes I_m)D_{PP_1}\triangle\bar{X} + h((b^T\mathcal{A}^{-1} \otimes I_m)D_{\gamma PP_1'} + \\
&\quad O(h))D_P\triangle\bar{X} + hPP_1(t_n)\delta_{n,s+1}, && (3.6a)\\
\triangle PQ_1x_n &= \rho\triangle PQ_1x_{n-1} + h\rho(PQ_1'(t_{n-1}) + O(h))\triangle Px_{n-1} + \\
&\quad (b^T\mathcal{A}^{-1} \otimes I_m)D_{PQ_1}\triangle\bar{X} + h((b^T\mathcal{A}^{-1} \otimes I_m)D_{\gamma PQ_1'} + O(h))D_P\triangle\bar{X} + \\
&\quad hPQ_1(t_n)\delta_{n,s+1}. && (3.6b)\\
\triangle Qx_n &= \rho\triangle Qx_{n-1} + (b^T\mathcal{A}^{-1} \otimes I_m)D_Q\triangle\bar{X} + hQ\delta_{n,s+1}. && (3.6c)
\end{aligned}$$

Inserting (3.3a) and (3.3b) into (3.6a) and (3.6b) and using the Taylor expansion at t_{n-1} provides again

$$\begin{aligned}
\|\triangle PP_1x_n\| &\leq \|\rho\triangle PP_1x_{n-1} + (b^T\mathcal{A}^{-1} \otimes I_m)(1\!\!1_s \otimes \triangle PP_1x_{n-1})\| + hC_1^{(n)}(\|\triangle Px_{n-1}\| + \delta)\\
&= \|(1 - b^T\mathcal{A}^{-1}1\!\!1_s)\triangle PP_1x_{n-1} + b^T\mathcal{A}^{-1}1\!\!1_s\triangle PP_1x_{n-1}\| + hC_1^{(n)}(\|\triangle Px_{n-1}\| + \delta)\\
&= \|\triangle PP_1x_{n-1}\| + hC^{(n)}(\|\triangle Px_{n-1}\| + \delta). && (3.7a)\\
\|\triangle PQ_1x_n\| &\leq |\,\rho\,|\ \|\triangle PQ_1x_{n-1}\| + \|(b^T\mathcal{A}^{-1} \otimes I_m)D_{PQ_1G_2^{-1}\delta_n}1\!\!1_s\| + hC_1^{(n)}(\|\triangle Px_{n-1}\| + \delta)\\
&\leq |\,\rho\,|\ \|\triangle PQ_1x_{n-1}\| + hC^{(n)}(\|\triangle Px_{n-1}\| + \frac{\delta}{h}), && (3.7b)
\end{aligned}$$

where $C^{(n)}$ and $C_1^{(n)}$ are constants.
Applying Lemma 3.1 to the recursions (3.7a-b) with $\|\triangle PP_1x_k\|$, $\|\triangle PQ_1x_k\|$, δ, and $\max_{i=1}^{n} C^{(i)}$ instead of y_k, z_k, D_j, C $(k = n-1, n,\ j = 1, 2)$ we obtain

$$\begin{aligned}
\|\triangle PP_1x_n\| &\leq C_1(\|\triangle PP_1x_0\| + h\|\triangle PP_1x_0\| + h\|\triangle PQ_1x_0\| + \delta) \text{ if } |\,\rho\,| \leq 1, && (3.8a)\\
\|\triangle PQ_1x_n\| &\leq C_1(h\|\triangle PP_1x_0\| + (|\,\rho\,|^n + h)\|\triangle PQ_1x_0\| + \delta) \text{ if } |\,\rho\,| < 1, && (3.8b)\\
\|\triangle PQ_1x_n\| &\leq C_1(h\|\triangle PP_1x_0\| + (1 + h)\|\triangle PQ_1x_0\| + \frac{\delta}{h}) \text{ if } |\,\rho\,| = 1. && (3.8c)
\end{aligned}$$

Now let $\tilde{x}_0 - x_0 := \delta_{0,s+1}$, then

$$\|\triangle Px_n\| \leq \|\triangle PP_1x_n\| + \|\triangle PQ_1x_n\| \leq C\delta, \tag{3.9}$$

where C is a constant and $|\rho| < 1$.

Next we consider the Q-component. Inserting (3.3c) into (3.6c) we obtain

$$\|\triangle Q x_n\| \quad \leq \quad |\,\rho\,|\;\|\triangle Q x_{n-1}\| + \frac{C_1}{h}(\|\triangle P Q_1 x_{n-1}\| + h\|\triangle P x_{n-1}\| + \delta).$$

By (3.9),

$$\|\triangle Q x_n\| \leq |\,\rho\,|\|\triangle Q x_{n-1}\| + \frac{C_2}{h}\delta \leq |\,\rho\,|^n\;\|\triangle Q x_0\| + \frac{C_2}{h}\delta\sum_{i=0}^{n-1}|\rho|^i \leq \frac{C}{h}\delta \qquad (3.10)$$

with C being constant, i.e. the IRK method is weakly unstable on the Q-component. □

Remarks

1. In a similar manner one can show that if $|\rho| = 1$, then the P-component is unstable, too. In this case we can obtain only an estimate of the form

$$\|\triangle P x_n\| \leq \frac{C}{h}\delta.$$

   Furthermore, the following estimate remains for the Q-component

$$\|\triangle Q x_n\| \leq \|\triangle Q x_0\| + \frac{C}{h^2}\Big(h\|\triangle P P_1 x_0\| + \|\triangle P Q_1 x_0\| + \frac{\delta}{h}\Big).$$

2. In the next section we will consider the consequences of the instability of IRK methods with $\|\rho\| = 1$ on the order of convergence and we will note that the method loses one order of convergence. This loss of order can be avoided in case of Hessenberg index 2 systems provided the method is modified by a projection techniques. These so-called *projected IRK methods* were introduced in [1], see also [7].

Now we study the convergence properties of the IRK method (2.1). We derive estimates for the local error of an IRK method when applied to the DAE (1.1). Then we study error propagation and derive estimates for the global error. Before we can state the main result of this section, we need some definitions.

Definition 3.1

1. *The local error d_n of the method (2.1) is given by*

$$d_n = \rho x(t_{n-1}) + \sum_{j=1}^{s} b_j \sum_{l=1}^{s} \hat{a}_{jl} X_l - x(t_n).$$

   Here X_l is obtained from (2.1b) with $x_{n-1} = x(t_{n-1})$.

2. *The j-th internal local truncation error $\delta_j^{(n)}$ at t_n of an s-stage IRK method (2.1) is given by*

$$\delta_j^{(n)} \quad = \quad x(t_{n-1}) + h\sum_{l=1}^{s} a_{jl} x'(t_{nl}) - x(t_{nj}), \qquad (3.11a)$$

$$\delta_{s+1}^{(n)} \;=\; x(t_{n-1}) + h\sum_{l=1}^{s} b_l x'(t_{nl}) - x(t_n). \tag{3.11b}$$

3. *The algebraic conditions $B(p)$ and $C(q)$ (Butcher identities) of an s-stage IRK method are given by*

$$B(p) \;:\; \sum_{i=1}^{s} b_i c_i^{k-1} = \frac{1}{k},\ \textit{for}\ k = 1,2,\ldots,p,$$
$$C(q) \;:\; \sum_{j=1}^{s} a_{ij} c_j^{k-1} = \frac{c_i^k}{k},\ \textit{for}\ k = 1,2,\ldots,q\ \textit{and all}\ i.$$

4. *The internal stage order K_I of an s-stage IRK method is given by*

$$K_I \;=\; \max\{k : B(k)\ \textit{and}\ C(k)\}.$$

The numerical solution of the IRK method (2.1) satisfies (3.2b), where δ_{nj}, $j = 1,2,\ldots,s$, represent errors in the solution of the system (2.1b). The true solution $x(t)$ of (1.1) satisfies

$$x(t_n) = \rho x(t_{n-1}) + \sum_{j=1}^{s}\sum_{l=1}^{s} b_j \hat{a}_{jl}(x(t_{nl}) + \delta_l^{(n)}) - \delta_{s+1}^{(n)}, \tag{3.12a}$$

$$\frac{1}{h}A(t_{nj})\sum_{l=1}^{s}\hat{a}_{jl}(x(t_{nl}) - x(t_{n-1}) + \delta_j^{(n)}) + B(t_{nj})x(t_{nj}) = q(t_{nj}), \tag{3.12b}$$

where $\delta_j^{(n)}$, $j = 1,2,\ldots,s+1$, are the internal local truncation errors at t_n.

Subtracting (3.12) from (3.2), we obtain

$$e_n = \rho e_{n-1} + (b^T\mathcal{A}^{-1}\otimes I_m)(\bar{E} - \bar{\delta}^{(n)}) + \delta_{s+1}^{(n)}, \tag{3.13a}$$

$$\frac{1}{h}A(t_{nj})\sum_{l=1}^{s}\hat{a}_{jl}E_l + B(t_{nj})E_j = \frac{1}{h}A(t_{nj})\sum_{l=1}^{s}\hat{a}_{jl}(\delta_l^{(n)} + e_{n-1}),\quad j = 1,2,\ldots,s. \tag{3.13b}$$

Here, $e_k := x_k - x(t_k)$, $k = n-1, n$, $E_j := X_j - x(t_{nj})$, $j = 1,2,\ldots,s$, $\bar{E} = (E_1^T, E_2^T, \ldots, E_s^T)^T \in I\!R^{ms}$, $\bar{\delta}^{(n)} = ((\delta_1^{(n)})^T, (\delta_2^{(n)})^T, \ldots, (\delta_s^{(n)})^T)^T \in I\!R^{ms}$, $\bar{e}_{n-1} = (e_{n-1}^T, e_{n-1}^T, \ldots, e_{n-1}^T)^T \in I\!R^{ms}$, and $\bar{\delta}_n = (\delta_{n1}^T, \delta_{n2}^T, \ldots, \delta_{ns}^T)^T \in I\!R^{ms}$.

According to Theorem 2.1 the system (3.13b) is, for sufficiently small h and $P' = 0$, uniquely solvable and the solution $\bar{E}$ satisfies (see (3.1))

$$D_{PP_1}\bar{E} \;=\; (D_{PP_1} + O(h))D_P(\bar{\delta}^{(n)} + \bar{e}_{n-1}) + O(h\|\bar{\delta}_n\|), \tag{3.14a}$$
$$D_{PQ_1}\bar{E} \;=\; D_{PQ_1G_2^{-1}}\bar{\delta}_n, \tag{3.14b}$$
$$\begin{aligned} D_Q\bar{E} \;=\; & \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QP_1P} + ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'}(D_{PP_1} - I_{ms}) - \\ & D_{QP_1G_2^{-1}BPP_1} + O(h)\Big]D_P(\bar{\delta}^{(n)} + \bar{e}_{n-1}) + O\Big(\frac{1}{h}\|\bar{\delta}_n\|\Big). \end{aligned} \tag{3.14c}$$

Multiplying (3.13a) by $PP_1(t_n)$, $PQ_1(t_n)$ and Q, and using the Taylor expansion at t_{n-1} on the right-hand side of the equations, we obtain

$$\begin{aligned} PP_1e_n &= \rho PP_1e_{n-1} + h\rho(PP_1'(t_{n-1}) + O(h))Pe_{n-1} + (b^T\mathcal{A}^{-1}\otimes I_m)D_{PP_1}(\bar{E} - \bar{\delta}^{(n)}) + \\ & \quad h\Big[((b^T\mathcal{A}^{-1}\otimes I_m)D_{\gamma PP_1'} + O(h))\Big]D_P(\bar{E} - \bar{\delta}^{(n)}) + PP_1(t_n)\delta_{s+1}^{(n)}, \qquad (3.15a) \\ PQ_1e_n &= \rho PQ_1e_{n-1} + h\rho(PQ_1'(t_{n-1}) + O(h))Pe_{n-1} + \\ & \quad (b^T\mathcal{A}^{-1}\otimes I_m)D_{PQ_1}\bar{E} + h\Big[((b^T\mathcal{A}^{-1}\otimes I_m)D_{\gamma PQ_1'} + O(h))\Big]D_P\bar{E} - \\ & \quad PQ_1(t_n)[(b^T\mathcal{A}^{-1}\otimes I_m)\delta^{(n)} - \delta_{s+1}^{(n)}], \qquad (3.15b) \\ Qe_n &= \rho Qe_{n-1} + (b^T\mathcal{A}^{-1}\otimes I_m)D_Q(\bar{E} - \bar{\delta}^{(n)}) + Q\delta_{s+1}^{(n)}\,, \qquad (3.15c) \end{aligned}$$

where $PP_1e_k := PP_1(t_k)e_k$, $PQ_1e_k := PQ_1(t_k)e_k$, $k = n-1, n$. Inserting (3.14) into (3.15) yields

$$PP_1e_n = PP_1e_{n-1} + O(h(\|Pe_{n-1}\| + \|\bar{\delta}^{(n)}\| + \|\bar{\delta}_n\|)) + PP_1(t_n)\delta_{s+1}^{(n)}\,, \qquad (3.16a)$$

$$\begin{aligned} PQ_1e_n &= \rho PQ_1e_{n-1} + O(h\|Pe_{n-1}\| + h\|\bar{\delta}^{(n)}\| + \|\bar{\delta}_n\|) - \\ & \quad PQ_1(t_n)[(b^T\mathcal{A}^{-1}\otimes I_m)\bar{\delta}^{(n)} - \delta_{s+1}^{(n)}], \qquad (3.16b) \end{aligned}$$

$$\begin{aligned} Qe_n &= \rho Qe_{n-1} + (b^T\mathcal{A}^{-1}\otimes I_m)\Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QP_1P} - D_{QP_1G_2^{-1}BPP_1} + \\ & \quad ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{QQ_1'}(D_{PP_1} - I_m) + O(h)\Big]D_P(\bar{\delta}^{(n)} + \bar{e}_{n-1}) - \\ & \quad (b^T\mathcal{A}^{-1}\otimes I_m)\bar{\delta}^{(n)} + Q\delta_{s+1}^{(n)} + O\Big(\frac{1}{h}\|\bar{\delta}_n\|\Big). \qquad (3.16c) \end{aligned}$$

The local error of the IRK method (2.1) is given by (3.16) with $x_{n-1} = x(t_{n-1})$, i.e. $e_{n-1} = 0$.

Theorem 3.2 *Let the assumptions of Theorem 2.1 be satisfied, and furthermore, $P' = 0$. Suppose that the IRK method satisfies the Butcher identities $B(p)$ and $C(q)$ for $p \geq q \geq 1$, and the numerical errors in the solution of the system (2.1b) are $\delta_{nj} = O(h^{q+\alpha})$. Then the local error satisfies:*

(a) $Pd_n = O(h^{q+1})$, $Qd_n = O(h^q)$ if $\alpha = 1$.

(b) $PP_1(t_n)d_n = O(h^{q+2})$, for $p \geq q+1$ and $\alpha = 1$.

(c) $PQ_1(t_n)d_n = O(h^{q+2})$,
if $\alpha = 2$ and the coefficients of the IRK method satisfy the identity

$$b^T\mathcal{A}^{-1}c^{q+1} = 1.$$

Here, $c^{q+1} = (c_1^{q+1}, c_2^{q+1}, \ldots, c_s^{q+1})^T$.

(d) $Pd_n = O(h^{q+2})$, if $\alpha = 2$, $b^T\mathcal{A}^{-1}c^{q+1} = 1$ and $p \geq q+1$.

(e) $Qd_n = O(h^{q+1})$,
if $\alpha = 2$ and the coefficients of the IRK method satisfy the identity

$$b^T\mathcal{A}^{-2}c^{q+1} = q+1.$$

Proof

(a) Follows directly from (3.16) with $e_{n-1} = 0$, $\bar{\delta}^{(n)} = O(h^{q+1})$, $\delta^{(n)}_{s+1} = O(h^{q+1})$, $\bar{\delta}_n = O(h^{q+1})$ and $Pe_n = PP_1e_n + PQ_1e_n$.

(b) Follows from (3.16a) with $e_{n-1} = 0$, $\bar{\delta}^{(n)} = O(h^{q+1})$, $\bar{\delta}_n = O(h^{q+1})$ and $\delta^{(n)}_{s+1} = O(h^{q+2})$ (because of $B(q+1)$).

(c) Inserting $e_{n-1} = 0$, $\bar{\delta}^{(n)} = O(h^{q+1})$ and $\bar{\delta}_n = O(h^{q+2})$ into (3.16b), we obtain

$$PQ_1(t_n)d_n = -PQ_1(t_n)[(b^T\mathcal{A}^{-1} \otimes I_m)\bar{\delta}^{(n)} - \delta^{(n)}_{s+1}] + O(h^{q+2}). \qquad (3.17)$$

For $q \leq p$, the following relations are valid

$$\begin{aligned}\bar{\delta}^{(n)} \quad &= \quad h^{q+1}\left[(\mathcal{A}c^q \otimes x^{(q+1)}(t_{n-1})) - \frac{1}{q+1}(c^{q+1} \otimes x^{(q+1)}(t_{n-1}))\right] + \\ &\quad\; O(h^{q+2}) \qquad (3.18a)\end{aligned}$$

and

$$\delta^n_{s+1} = h^{q+1}\left(b^Tc^q - \frac{1}{q+1}\right)x^{(q+1)}(t_{n-1}) + O(h^{q+2}). \qquad (3.18b)$$

This yields

$$\begin{aligned}(b^T\mathcal{A}^{-1} \otimes I_m)\bar{\delta}^{(n)} - \delta^{(n)}_{s+1} \quad &= \quad h^{q+1}(b^T\mathcal{A}^{-1} \otimes I_m)\Big[(\mathcal{A}c^q) \otimes x^{(q+1)}(t_{n-1}) - \\ \frac{1}{q+1}(c^{q+1} \otimes x^{(q+1)}(t_{n-1}))\Big] \quad &- \quad h^{q+1}\left(b^Tc^q - \frac{1}{q+1}\right)x^{(q+1)}(t_{n-1}) + O(h^{q+2}) \\ &= \quad \frac{h^{q+1}}{q+1}(1 - b^T\mathcal{A}^{-1}c^{q+1})x^{(q+1)}(t_{n-1}) + O(h^{q+2}).\end{aligned}$$

Inserting this equation into (3.17) and using $b^T\mathcal{A}^{-1}c^{q+1} = 1$ we obtain

$$PQ_1(t_n)d_n = O(h^{q+2}).$$

(d) Follows directly from (b) and (c).

(e) Inserting (3.18a) into (3.16c) and using $\bar{\delta}^{(n)} = O(h^{q+1})$, $\delta^{(n)}_{s+1} = O(h^{q+1})$, $\bar{\delta}_n = O(h^{q+2})$ and $e_{n-1} = 0$ we obtain

$$Qd_n \quad = \quad h^qQP_1P(t_n)((b^T\mathcal{A}^{-1}c^q - \frac{1}{q+1}b^T\mathcal{A}^{-2}c^{q+1}) \otimes x^{(q+1)}(t_{n-1})) + O(h^{q+1}).$$

Inserting the identity $b^T\mathcal{A}^{-2}c^{q+1} = q+1$ into this equation and using $B(q)$ and $C(q)$ we have

$$Qd_n = O(h^{q+1}) \ . \ \square$$

Theorem 3.2 gives conditions on the coefficients of an IRK method so that it attains a given order of the local error when applied to the DAE (1.1) with constant nullspace. The conditions $b^T\mathcal{A}^{-1}c^{q+1} = 1$ and $b^T\mathcal{A}^{-2}c^{q+1} = q+1$ have been studied in [2] for IRK methods applied to linear constant coefficient systems of index 1 and 2. Furthermore,

Brennan, Campbell and Petzold [2] derive conditions for the local error of an IRK method applied to linear constant coefficient systems of arbitrary index k.
Now we use the recursions (3.16) in order to derive estimates for the global error.

Theorem 3.3 *Let the assumptions of Theorem 2.1 be satisfied and, furthermore, $P' = 0$. Suppose the IRK method satisfies $|\rho| < 1$. For $q \geq 1$, let $e_0 = O(h^{q+\alpha})$ be the errors in the initial conditions, and $\delta_{ij} = O(h^{q+\alpha})$ the numerical errors in the solution of the system (2.1b) at the i-th integration step. Then the global error e_n satisfies:*

- $\|Pe_n\| \leq Ch^q$ *if* $B(q)$, $C(q)$ *and* $\alpha = 0$,
- $\|Pe_n\| \leq Ch^{q+1}$ *if* $B(q+1)$, $C(q)$ *and* $\alpha = 1$,
- $\|Qe_n\| \leq Ch^q$ *if* $B(q)$, $C(q)$ *and* $\alpha = 1$,
- $\|Qe_n\| \leq Ch^{q+1}$ *if* $B(q+1)$, $C(q)$, $b^T\mathcal{A}^{-1}c^{q+1} = 1$, $b^T\mathcal{A}^{-2}c^{q+1} = q+1$ *and* $\alpha = 2$.

Proof From (3.16a-b) we obtain inmediately the estimates

$$\begin{aligned} \|PP_1e_n\| &\leq \|PP_1e_{n-1}\| + hC^{(n)}(\|PP_1e_{n-1}\| + \|PQ_1e_{n-1}\| + D_1), &(3.19a)\\ \|PQ_1e_n\| &\leq |\,\rho\,|\,\|PQ_1e_{n-1}\| + hC^{(n)}(\|PP_1e_{n-1}\| + \|PQ_1e_{n-1}\| + \frac{D_2}{h})\,, &(3.19b)\end{aligned}$$

with some constants $C^{(n)}$ and

$$D_1 = \begin{cases} O(h^q) \text{ if } \alpha = 0,\ B(q) \text{ and } C(q), \\ O(h^{q+1}) \text{ if } \alpha = 1,\ B(q+1) \text{ and } C(q), \end{cases} \tag{3.20a}$$

$$D_2 = \begin{cases} O(h^q) \text{ if } \alpha = 0,\ B(q) \text{ and } C(q), \\ O(h^{q+1}) \text{ if } \alpha = 1,\ B(q) \text{ and } C(q), \\ O(h^{q+2}) \text{ if } \alpha = 2,\ b^T\mathcal{A}^{-1}c^{q+1} = 1,\ B(q) \text{ and } C(q). \end{cases} \tag{3.20b}$$

Applying Lemma 3.1 to the recursions (3.19) we obtain

$$\begin{aligned} \|PP_1e_n\| &\leq C_1(\|PP_1e_0\| + h\|PP_1e_0\| + h\|PQ_1e_0\| + D_1 + D_2), &(3.21a)\\ \|PQ_1e_n\| &\leq C_1(h\|PP_1e_0\| + (|\rho|^n + h)\|PQ_1e_0\| + hD_1 + D_2). &(3.21b)\end{aligned}$$

Now the estimates for Pe_n follow from (3.20), (3.21) and $\|Pe_n\| \leq \|PP_1e_n\| + \|PQ_1e_n\|$. In order to prove the first estimate for Qe_n we consider (3.16c) and (3.21b). This last inequality together with (3.20) yields

$$PQ_1e_n = O(h^{q+1}), \text{ if } B(q),\ C(q) \text{ and } \alpha = 1.$$

Inserting this last equation together with $Pe_n = O(h^q)$, $\bar{\delta}_n = O(h^{q+1})$, $\bar{\delta}^{(n)} = O(h^{q+1})$ and $\delta^{(n)}_{s+1} = O(h^{q+1})$ into (3.16c), we obtain

$$Qe_n = \rho Qe_{n-1} + O(h^q).$$

Now, the first estimate for Qe_n follows directly from this recursion. In order to prove the second estimate for Qe_n, we insert (3.20) with $\alpha = 2$, $B(q+1)$, $C(q)$ and $b^T\mathcal{A}^{-1}c^{q+1} = 1$ into (3.21b). This implies

$$PQ_1e_n = O(h^{q+2}).$$

Inserting this equation together with $Pe_n = O(h^{q+1})$, $\bar{\delta}_n = O(h^{q+2})$, $\delta_{s+1}^{(n)} = O(h^{q+2})$ and $\bar{\delta}^{(n)} = O(h^{q+1})$ into (3.16c) we obtain

$$Qe_n = \rho Qe_{n-1} + \frac{1}{h}(b^T\mathcal{A}^{-2} \otimes I_m)D_{QP_1P}\bar{\delta}^{(n)} + O(h^{q+1}).$$

Using the identity $b^T\mathcal{A}^{-2}c^{q+1} = q+1$ we have

$$Qe_n = \rho e_{n-1} + O(h^{q+1}).$$

The second estimate for Qe_n is an immediate consequence. $\square$

For many important methods, for example the Gauss-Legendre methods, it holds $|\rho| = 1$. This means that the statements of Theorem 3.3 are not valid for this class of IRK. The following theorem shows how strong the order reduction is, when IRK methods with $|\rho| = 1$ are applied to linear DAEs with index 2 and constant nullspace.

Theorem 3.4 *Let the assumption of Theorem 2.1 be true. Suppose the IRK method satisfies $|\rho| = 1$ and, furthermore, assume that for $q \geq 1$, $e_0 = O(h^{q+\beta})$ is the error in the initial condition and $\delta_{ij} = O(h^{q+\alpha})$ is the numerical error in the solution of the system (2.1b) at the i-th integration step. Then the global error e_n satisfies:*

- $\|Pe_n\| \leq Ch^q$, $\|Qe_n\| \leq Ch^{q-2}$ *if* $\alpha = 1$, $\beta = 0$, $B(q)$ *and* $C(q)$ *for* $q \geq 2$,
- $\|Pe_n\| \leq Ch^{q+1}$, $\|Qe_n\| \leq Ch^{q-1}$ *if* $\alpha = 2$, $\beta = 1$, $B(q+1)$, $C(q)$ *and* $A_1(q+1)$,
- $\|Pe_n\| \leq Ch^{q+1}$, $\|Qe_n\| \leq Ch^{q-1}$ *if* $\alpha = 2$, $\beta = 1$, $B(q+1)$, $C(q)$ *and* $\rho = -1$.

Proof The first and the second estimate for Pe_n and Qe_n follow directly from the assumptions of the theorem and by application of Lemma 3.1 to the recursions (3.20) with $|\rho| = 1$. In order to prove the last estimates we consider the equation (3.16b) or, equivalently,

$$PQ_1(t_n)e_n = \rho(I_m + hPQ_1'(t_{n-1}))PQ_1(t_{n-1})e_{n-1} + A_n + B_n, \tag{3.22}$$

$$\begin{aligned}
\text{with} \quad A_n \;&:=\; h\rho PQ_1'PP_1(t_{n-1})e_{n-1} + O(h^2\|Pe_{n-1}\|) - \\
&\quad\; h\Big[((b^T\mathcal{A}^{-1} \otimes I_m)D_{\gamma PQ_1'} + O(h))\Big](D_{PP_1} + O(h))D_P(\bar{\delta}^{(n)} + \bar{e}_{n-1}) - \\
&\quad\; h\Big[((b^T\mathcal{A}^{-1} \otimes I_m)D_{\gamma PQ_1'} + O(h))\Big]D_P\bar{\delta}^{(n)} + PQ_1(t_n)\delta_{s+1}^{(n)} + O(h\|\bar{\delta}_n\|), \\
B_n \;&:=\; -\sum_{j=1}^{s}\sum_{l=1}^{s} b_j\hat{a}_{jl}PQ_1\delta_l^{(n)}.
\end{aligned}$$

From the assumptions of the theorem we have $A_n = O(h^{q+2})$. Furthermore, the recursion (3.22) together with $\rho = -1$ yields

$$PQ_1(t_n)e_n = (-1)^n\Pi_0 PQ_1(t_0)e_0 + \sum_{i=1}^{n}(-1)^{n-i}\Pi_i A_i + \sum_{i=1}^{n}(-1)^{n-i}\Pi_i B_i$$

where $\Pi_i := \prod_{j=1}^{n-i}(I_m + hPQ_1'(t_{n-j}))$ for $i = 0, 1, \ldots, n$, i. e. $\Pi_i = O(1)$.

Expanding the latter term of the right-hand side of this equation into a Taylor series and grouping the respective terms in the so obtained sums together two at a time, and then bounding the resulting sums, we obtain

$$PQ_1e_n = O(h^{q+1}).$$

Note that we gain one power of h in this process because of the alternating signs in the Taylor series. For the PP_1-component (3.21a), $B(q+1)$, $C(q)$, $\alpha = \beta = 1$ and $|\rho| = 1$ imply

$$PP_1e_n = O(h^{q+1}).$$

The last estimates for PQ_1e_n and PP_1e_n together with $\|Pe_n\| \leq \|PP_1e_n\| + \|PQ_1e_n\|$ yields

$$\|Pe_n\| \leq Ch^{q+1}.$$

The estimate for Qe_n follows then directly from (3.16c). $\square$

Lobatto IIIA methods

For these methods the coefficient matrix $\mathcal{A}$ is singular, because in this case $\mathcal{A}$ can be expressed in the form

$$\mathcal{A} = \begin{pmatrix} 0 & 0 \\ \mathbf{a} & \underline{\mathcal{A}} \end{pmatrix}$$

with $\mathbf{a} = (a_{21}, a_{31}, \dots, a_{s1})^T \in \mathbb{R}^{s-1}$, $\underline{\mathcal{A}} \in L(\mathbb{R}^{s-1})$ and $\underline{\mathcal{A}}$ nonsingular.

The best known of such methods is the implicit trapezoidal method. To implement these methods for the DAE (1.1) we assume that initial values for the derivates of all the variables are given. So we set in the first step $X'_{01} := x'_0$ and at the end of each step we set $X'_{n1} = X'_{n-1,s}$. Since $X_1 = x_{n-1}$, we need to solve (2.1b) only for $\hat{X} = (X_2^T, X_3^T, \dots, X_s^T)^T \in \mathbb{R}^{m(s-1)}$ in the following system

$$\begin{aligned}[\hat{D}_A(\underline{\mathcal{A}}^{-1} \otimes I_m) + h\hat{D}_B]\hat{X} &= h\hat{D}_q \mathbb{1}_{s-1} + \hat{D}_A(\underline{\mathcal{A}}^{-1} \otimes I_m)(\mathbb{1}_{s-1} \otimes x_{n-1}) + \\ &\quad h\hat{D}_A(\underline{\mathcal{A}}^{-1} \otimes I_m)(\mathbf{a} \otimes X'_1).\end{aligned}$$

Here and subsequently in this section we define, for a matrix $N(t) \in L(\mathbb{R}^m)$, $\hat{D}_N := \mathrm{diag}(N(t_{n2}), N(t_{n3}), \dots, N(t_{ns})) \in L(\mathbb{R}^{m(s-1)})$.

According to Theorem 2.1 this system is uniquely solvable. The main differnce to IRK methods with nonsingular $\mathcal{A}$ is that now the internal stages X_j, $j = 1, 2, \dots, s$ also depend on the Q-component of x_{n-1}. This fact makes the study of stability and convergence on this class of IRK methods more complicate. In the case of DAEs (1.1) with constant matrix A we study here the stability and convergence properties of the method. Then instead of (3.7) we have the estimates

$$\begin{aligned}\|\triangle PP_1x_n\| &\leq \|\triangle PP_1x_{n-1}\| + hC^{(n)}(\|\triangle PP_1x_{n-1}\| + h\|\triangle Qx_{n-1}\| + \delta), &(3.23a)\\ \|\triangle PQ_1x_n\| &\leq \delta, &(3.23b)\end{aligned}$$

and for the Q-component we have

$$h\|\triangle Qx_n\| \leq h\|\triangle Qx_{n-1}\| + hC^{(n)}(\|\triangle PP_1x_{n-1}\| + h\|\triangle Qx_{n-1}\| + \frac{\delta}{h}). \quad (3.23c)$$

Applying Lemma 3.1 to the recursions (3.23a) and (3.23c) we obtain

$$\|\triangle PP_1x_n\| \le C\delta, \quad h\|\triangle Qx_n\| \le C\frac{\delta}{h}.$$

These estimates together with (3.23b) show that Lobatto IIIA methods applied to the DAE (1.1) with $P' = 0$ are weakly unstable on the Q-component but stable on the P-component.

Instead of (3.16) we have for the global error e_n

$$\begin{aligned} \|PP_1e_n\| &\le \|PP_1e_{n-1}\| + hC^{(n)}(\|PP_1e_{n-1}\| + h\|Qe_{n-1}\| + D_1), & (3.24a)\\ \|PQ_1e_n\| &\le C\delta, & (3.24b)\\ h\|Qe_n\| &\le h\|Qe_{n-1}\| + hC^{(n)}\Big(\|PP_1e_{n-1}\| + h\|Qe_{n-1}\| + \frac{D_2}{h}\Big), & (3.24c)\end{aligned}$$

with

$$\begin{aligned} D_1 &= \begin{cases} O(h^2) & \text{if } s=2 \text{ and } \delta = O(h^2),\\ O(h^{s+1}) & \text{if } s\ge 3 \text{ and } \delta = O(h^{s+1}),\end{cases}\\ D_2 &= \begin{cases} O(h^2) & \text{if } s=2 \text{ and } \delta = O(h^2),\\ O(h^{s+1}) & \text{if } s\ge 2 \text{ and } \delta = O(h^{s+1}).\end{cases}\end{aligned}$$

Applying Lemma 3.1 to the recursions (3.24a) and (3.24c) we obtain

$$\begin{aligned} \|PP_1e_n\| &\le \begin{cases} Ch^2 & \text{if } s=2 \text{ and } \delta = O(h^2),\\ Ch^{s+1} & \text{if } s\ge 3 \text{ and } \delta = O(h^{s+1}),\end{cases}\\ \|Qe_n\| &\le Ch^{s-1} \quad \text{if } \delta = O(h^{s+1}).\end{aligned}$$

If $\delta = O(h^{s+2})$ and $s > 2$ is an even number, we can write (3.24c) in the following form

$$Qe_n = -(I_m + O(h))Qe_n + A_n + B_n$$

with $\|A_n\| = O(h^{s+1})$ and $B_n = -\sum_{l=2}^{s} \hat{a}_{sl}QQ_1(t_{n-1})\delta_l^{(n)}$.

So we can obtain in this case (analogously to the proof of the last statement of Theorem 3.4) the following estimate for Qe_n

$$\|Qe_n\| \le Ch^s.$$

Remarks

1. For nonlinear DAEs with index 2 in Hessenberg form Kværnø [9] has shown that the order of accuracy for some stiffly accurate formulas in the differential component is the same as the nonstiff order of accuracy. In particular, the Radau IIA and Lobatto IIIC methods retain their nonstiff order of accuracy in the differential component. The analysis makes use of the theory of Butcher series and rooted trees and exploid the Hessenberg structure of the DAE. The same order results based on a similar analysis are given by [5] for the same class of DAEs. These

earlier results make clear that, for special IRK methods, Theorem 3.2 does not give an optimal estimate for the local error of the P-component. These particularly well suited IRK methods for the numerical solution of DAEs are those IRKs for which $c_s = 1$ and $b_j = a_{sj}$, $j = 1, 2, \ldots, s$. Following Griepentrog and März [4] this class of IRK methods is called IRK(DAEs). It is expected that similar results hold in the situation considered in the present section. Note that $\rho = 0$ for IRK(DAEs).

2. Provided the IRK (2.1) is feasible, we can obtain the equation (2.6). Solving it with respect to $D_{PP_1}\bar{X}$, $D_{PQ_1}\bar{X}$ and $D_Q\bar{X}$ provides

$$
\begin{aligned}
D_{PP_1}\bar{X} &= (I_{ms} + O(h))[h(\mathcal{A} \otimes I_m)D_{PP_1G_2^{-1}q}1\!\!1_s + D_{PP_1}(1\!\!1_s \otimes x_{n-1})] + \\
&\quad O(\|D_q 1\!\!1_s\|) + O(\|D_{QQ_1}(1\!\!1_s \otimes x_{n-1})\|) + O(h\|x_{n-1}\|), && (3.25a)\\
D_{PQ_1}\bar{X} &= D_{PQG_2^{-1}q}1\!\!1_s, && (3.25b)\\
D_Q\bar{X} &= (\mathcal{D}^{-1} + O(h))\Big[\frac{1}{h}(\mathcal{A}^{-1} \otimes I_m)D_{QQ_1G_2^{-1}q}1\!\!1_s + \\
&\quad D_{QP_1G_2^{-1}q}1\!\!1_s + ((\mathcal{C} \odot \mathcal{A}^{-1}) \otimes I_m)D_{(QQ_1)'PQ_1G_2^{-1}q}1\!\!1_s + \\
&\quad O(h\|D_{PP_1G_2^{-1}q}1\!\!1_s\|) - \frac{1}{h}(\mathcal{A}^{-1} \otimes I_m)D_{QQ_1}(1\!\!1_s \otimes x_{n-1}) - \\
&\quad ((\mathcal{C} \odot \mathcal{A}^{-1}) \otimes I_m)D_{(QQ_1)'}(1\!\!1_s \otimes x_{n-1}) - D_{QP_1G_2^{-1}BPP_1}(1\!\!1_s \otimes x_{n-1}) + \\
&\quad ((\mathcal{C} \odot \mathcal{A}^{-1}) \otimes I_m)D_{QQ_1'PP_1}(1\!\!1_s \otimes x_{n-1}) + O(h\|x_{n-1}\|)\Big]. && (3.25c)
\end{aligned}
$$

(3.25) is essentially different from (3.1), because now the internal stages X_j, j = 1,2,...,s, also depend on the Q-component of the approximation x_{n-1}. That means that the instability of the Q-component can lead to instability of the P-component and the weak instability can be transformed into exponential instability. For example, applying the implicit Euler method to the DAE of Example 2.2 we obtain the recursions

$$
\begin{aligned}
x_n^{(1)} &= q_n^{(1)} - \eta t_n x_n^{(2)},\\
x_n^{(2)} &= \frac{\eta}{1+\eta}x_{n-1}^{(2)} + \frac{1}{1+\eta}(q_n^{(2)} - \frac{1}{h}(q_n^{(1)} - q_{n-1}^{(1)})), \qquad \eta \neq 1.
\end{aligned}
$$

Evidently, this recursion has an exponentially unstable behaviour when $\eta < -1/2$. More generally, one can derive conditions on the method and the problem such that the recursion is weakly unstable. However, these conditions are so restrictive that the conclusion must be: The IRK (1.2) *cannot* be applied to linear DAEs with *variable* nullspace. More details can be found in [8].

4. A modified IRK method for the numerical solution of DAEs with index 2 and variable nullspace

To overcome the drawbacks of IRK methods in the variable nullspace case we propose a new realization of the IRK methods for the DAE (1.1). The new method seems to be promising for solving the DAE (1.1) with index 2 and variable nullspace.

The most important difficulties in the numerical integration of the DAE (1.1) with variable nullspace by IRK methods lie in the fact that the P-component, which is stable, also depends on the Q-component of the numerical approximation x_{n-1}. For this reason we now first split the DAE into the two components Px and Qx, and in each one we apply the IRK method (1.2) separately. Let us consider the DAE (1.1) with nonconstant nullspace $N(t)$. Under the assumption that (1.1) is transferable with index 2 we can decouple it in the following way:

$$[A(t) + (B(t) - A(t)P'(t))Q(t)][P(t)(P(t)x(t))' + Q(t)x(t)] + B(t)P(t)x(t) = q(t).$$

Obviously, this equation is equivalent to (1.1). Now we set $y(t) := P(t)x(t)$ and $z(t) := Q(t)x(t)$, so the DAE (1.1) can be written as

$$A(t)y'(t) + B(t)y(t) + A_1(t)z(t) = q(t). \tag{4.1a}$$

The new variables y and z satisfy the identities $P(t)y(t) = y(t)$ and $Q(t)z(t) = z(t)$. We combine these equations to $y(t) - P(t)y(t) + z(t) - Q(t)z(t) = 0$, or, equivalently:

$$Q(t)y(t) + P(t)z(t) = 0. \tag{4.1b}$$

Note that (4.1) is again an index 2 system.

Formulation of the method
To solve the IVP (1.1) we compute the approximations y_{nj} and z_{nj} of the solution in t_{nj} in the way described below and then we construct the approximation x_{nj} of $x(t_{nj})$ via the formula

$$x_{nj} := y_{nj} + z_{nj}\,, \qquad j = 1, 2, \ldots, s.$$

Note that the exact solution of (1.1) satisfies $x(t) = y(t) + z(t)$.

Numerical approach for the y-component

In the first step we set $y_0 := P(t_0)x_0$ and in the n-th integration step we construct the approximation $y_n = y_{n0}$ via the formulas

$$y_n = \rho y_{n-1} + \sum_{j=1}^{s}\sum_{l=1}^{s} b_j \hat{a}_{jl} Y_l, \tag{4.2}$$

where Y_l is defined by

$$\begin{aligned} A(t_{nj})\sum_{l=1}^{s}\hat{a}_{jl}Y_l + hB(t_{nj})Y_j + hA_1(t_{nj})Z_j &= hq(t_{nj}) + A(t_{nj})\sum_{l=1}^{s}\hat{a}_{jl}y_{n-1}\,, & (4.3\text{a})\\ Q(t_{nj})Y_j + P(t_{nj})Z_j &= 0. & (4.3\text{b}) \end{aligned}$$

In the n-th integration step we set $y_{nj} = Y_j$, for $j = 1, 2, \ldots, s$.

Numerical approach for the z-component

If the system (4.3) is uniquely solvable with respect to (Y_j, Z_j) (we prove this statement below), we obtain a numerical solution for the z-component at t_{nj} directly from the internal stages Z_j, i. e. we can take Z_j as an approximation of $z(t_{nj})$ in the n-th integration step. Furthermore, it is possible to obtain an approximation for the z-component at t_i, $i = 1, 2, \ldots, n$, via the formula

$$z_n = \rho z_{n-1} + \sum_{j=1}^{s}\sum_{l=1}^{s} b_j \hat{a}_{jl} Z_l.$$

We renounce here to consider this last approximation, because the calculation of (Y_j, Z_j) from (4.3) does not depend on the approximation z_{n-1}.

Existence and uniqueness of the Runge-Kutta solution
Using the notation of Section 2, the system (4.3) becomes equivalent to

$$\begin{aligned} [D_A(\mathcal{A}^{-1}\otimes I_m) + hD_B]\bar{Y} + hD_{A_1}\bar{Z} &= hD_q 1\!\!1_s + D_A(\mathcal{A}^{-1}1\!\!1_s \otimes y_{n-1}), &(4.4a)\\ D_Q\bar{Y} + D_P\bar{Z} &= 0. &(4.4b)\end{aligned}$$

Here, $\bar{Y} = (Y_1^T, Y_2^T, \ldots, Y_s^T)^T$ and $\bar{Z} = (Z_1^T, Z_2^T, \ldots, T_s^T)^T$.

Together with (4.4) we consider the perturbed system

$$\begin{aligned} [D_A(\mathcal{A}^{-1}\otimes I_m) + hD_B]\bar{\tilde{Y}} + hD_{A_1}\bar{\tilde{Z}} &= hD_{(q+\delta_n)} 1\!\!1_s + D_A(\mathcal{A}^{-1}1\!\!1_s \otimes \tilde{y}_{n-1}), &(4.5a)\\ D_Q\bar{\tilde{Y}} + D_P\bar{\tilde{Z}} &= \bar{\Theta}, &(4.5b)\end{aligned}$$

where $\bar{\tilde{Y}} := (\tilde{Y}_1^T, \tilde{Y}_2^T, \ldots, \tilde{Y}_s^T)^T$, $\bar{\tilde{Z}} := (\tilde{Z}_1^T, \tilde{Z}_2^T, \ldots, \tilde{Z}_s^T)^T$, $\bar{\Theta} := (\theta_{n1}^T, \theta_{n2}^T, \ldots, \theta_{ns}^T)^T$, and we suppose that the perturbations satisfy $\|\delta_{ij}\| \leq \delta$ and $\|\theta_{ij}\| \leq \delta$ for all $i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, s$.
Obviously, (4.4) is equivalent to (4.5) if $\delta = 0$.

Theorem 4.1 *Suppose that the IVP (1.1) is tractable with index 2 and the Runge-Kutta matrix $\mathcal{A}$ is nonsingular, then the systems (4.4) and (4.5) are uniquely solvable with respect to $(\bar{Y}, \bar{Z})$ and $(\bar{\tilde{Y}}\bar{\tilde{Z}})$, respectively, for sufficiently small h.*

Proof Using the techniques of Section 2 and under consideration of (4.5b) we decouple the system (4.5a) into the PP_1-, Q_1- and Q-components. This yields:

$$\begin{aligned} [D_{PP_1}(\mathcal{A}^{-1}\otimes I_m) + hD_{PP_1G_2^{-1}BPP_1}]\bar{\tilde{Y}} &= hD_{PP_1G_2^{-1}(q+\delta_n)}1\!\!1_s + D_{PP_1}(\mathcal{A}^{-1}1\!\!1_s\otimes\tilde{y}_{n-1}) - \\ &\quad\ hD_{PP_1}\bar{\Theta}, &(4.6a)\\ D_{Q_1}\bar{\tilde{Y}} &= D_{Q_1G_2^{-1}(q+\delta_n)}1\!\!1_s\,, &(4.6b)\\ [-D_{QQ_1}(\mathcal{A}^{-1}\otimes I_m) + hD_{QP_1G_2^{-1}BPP_1}]\bar{\tilde{Y}} &+ hD_Q\bar{\tilde{Z}} = hD_{QP_1G_2^{-1}(q+\delta_n)}1\!\!1_s - \\ D_{QQ_1}(\mathcal{A}^{-1}1\!\!1_s\otimes\tilde{y}_{n-1}) &- hD_{QP_1}\bar{\Theta}. &(4.6c)\end{aligned}$$

Inserting (4.6b) into (4.6a) and using the Taylor expansions, we obtain

$$\begin{aligned} [(\mathcal{A}^{-1}\otimes I_m) + O(h)]D_{PP_1}\bar{\tilde{Y}} &= hD_{PP_1G_2^{-1}(q+\delta_n)}1\!\!1_s - R_{PP_1}D_{PQ_1G_2^{-1}(q+\delta_n)}1\!\!1_s + \\ &\quad\ D_{PP_1}(\mathcal{A}^{-1}1\!\!1_s\otimes\tilde{y}_{n-1}) + O(h\|\bar{\Theta}\|).\end{aligned}$$

Here R_{PP_1} is as in (2.3c).

The matrix on the left-hand side of this equation is invertible if h is sufficiently small. Therefore, we can write

$$\begin{aligned} D_{PP_1}\bar{\tilde{Y}} &= [(\mathcal{A}\otimes I_m)+O(h)][hD_{PP_1G_2^{-1}(q+\delta_n)}1\!\!1_s - R_{PP_1}D_{PQ_1G_2^{-1}(q+\delta_n)}1\!\!1_s + \\ & \quad D_{PP_1}(\mathcal{A}^{-1}1\!\!1_s\otimes\tilde{y}_{n-1})+O(h\|\bar{\Theta}\|)]. \end{aligned} \tag{4.7}$$

Using the equation $\bar{\tilde{Y}} = D_P\bar{\tilde{Y}} + D_Q\bar{\Theta} = D_{PP_1}\bar{\tilde{Y}} + D_{PQ_1}\bar{\tilde{Y}} + D_Q\bar{\Theta}$, together with (4.6b) and (4.7), we obtain $\bar{\tilde{Y}}$. $\bar{\tilde{Z}}$ follows directly from (4.6c) and the identity $D_P\bar{\tilde{Z}} = \bar{\tilde{Z}} - D_Q\bar{\tilde{Z}} = D_P\bar{\Theta}$. $\square$

Stability and convergence
In addition to the vector y_n obtained from (5.2) we consider

$$\tilde{y}_n = \rho\tilde{y}_{n-1} + \sum_{j=1}^{s}\sum_{l=1}^{s} b_j\hat{a}_{jl}\tilde{Y}_l + h\delta_{n,s+1}\,.$$

Subtracting this equation from (4.2) we obtain

$$\triangle y_n = \rho\triangle y_{n-1} + (b^T\mathcal{A}^{-1}\otimes I_m)\triangle\bar{Y} + h\delta_{n,s+1}\,. \tag{4.8}$$

Here $\triangle y_k = \tilde{y}_k - y_k$, $k = n-1, n$ and $\triangle\bar{Y} := \bar{\tilde{Y}} - \bar{Y}$.
Subtracting (4.4) from (4.5) and using the techniques of the feasibility proof on the so obtained equation, we have now

$$
\begin{aligned}
D_{PP_1}\triangle\bar{Y} &= h((\mathcal{A}\otimes I_m)+O(h))D_{PP_1G_2^{-1}\delta_n}1\!\!1_s - h[(\mathcal{A}(\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{(PP_1)'} + \\
&\quad O(h)]D_{PQ_1G_2^{-1}\delta_n}1\!\!1_s + (D_{PP_1}+O(h))(1\!\!1_s\otimes\triangle y_{n-1}) + O(h\|\bar{\Theta}\|), \quad (4.9a)\\
D_{PQ_1}\triangle\bar{Y} &= D_{PQ_1G_2^{-1}\delta_n}1\!\!1_s\,, \quad (4.9b)\\
\triangle\bar{Z} &= \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QQ_1} + D_{QP_1} + ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{(QQ_1)'PQ_1} + \\
&\quad O(h)\Big]D_{G_2^{-1}\delta_n}1\!\!1_s + \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{QP_1P} - D_{QP_1G_2^{-1}BPP_1} + \\
&\quad ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{(QQ_1)'}(D_{PP_1}-I_{ms}) + O(h)\Big](1\!\!1_s\otimes\triangle y_{n-1}) + \\
&\quad O(\|\bar{\Theta}\|). \quad (4.9c)
\end{aligned}
$$

Now we can use (4.8) and (4.9) to prove the stability and convergence statements. For this we will need the following lemma, too. For the proof we refer to [8].

Lemma 4.1 *Suppose that the sequences of positive numbers (y_n), (z_n) and (w_n) satisfy*

$$
\begin{aligned}
y_n &\le y_{n-1} + hC(y_{n-1}+z_{n-1}+w_{n-1}+D_1),\\
z_n &\le |\,\rho\,|z_{n-1} + hC(y_{n-1}+z_{n-1}+w_{n-1}+\frac{D_2}{h}),\\
w_n &\le |\rho|w_{n-1} + hC\Big(z_{n-1}+y_{n-1}+w_{n-1}+\frac{D_3}{h}\Big).
\end{aligned}
$$

Then we have for $hn \le c_0$:

$$
\begin{aligned}
y_n &\le C_1(y_0+hz_0+hw_0+D_1+D_2+D_3), \quad \textit{if } |\rho|\le 1,\\
z_n &\le C_1(hy_0+(|\rho|^n+h)z_0+hw_0+hD_1+D_2+hD_3), \quad \textit{if } |\rho|<1,\\
w_n &\le C_1(hy_0+hz_0+(|\rho|^n+h)w_0+hD_1+hD_2+D_3), \quad \textit{if } |\rho|<1.
\end{aligned}
$$

Furthermore

$$
\begin{aligned}
z_n &\le C_1(hy_0+z_0+h^2w_0+hD_1+\frac{D_2}{h}+D_3) \quad \textit{if } |\rho|=1,\\
w_n &\le C_1(hy_0+hz_0+hw_0+hD_1+D_2+\frac{D_3}{h}) \quad \textit{if } |\rho|=1.
\end{aligned}
$$

Theorem 4.2 *Let the assumption of Theorem 4.1 be satisfied. If the coefficients of the IRK method satisfy $|\rho|<1$, then the method (4.2), (4.3) is weakly unstable. This weak instability refers to the z-component of the numerical solution. The y-component is stable.*

Proof Multiplying (4.8) by $PP_1(t_n)$, $PQ_1(t_n)$ and $Q(t_n)$, and using the Taylor expansions at t_{n-1} on the right-hand side of the resulting equations we obtain

$$
\begin{aligned}
\triangle PP_1 y_n &= \rho \triangle PP_1 y_{n-1} + h\rho((PP_1)'(t_{n-1}) + O(h))\triangle y_{n-1} + \\
&\quad (b^T\mathcal{A}^{-1} \otimes I_m) D_{PP_1} \triangle \bar{Y} + h((b^T\mathcal{A}^{-1} \otimes I_m) D_{\gamma(PP_1)'} + O(h))\triangle \bar{Y} + \\
&\quad hPP_1(t_n)\delta_{n,s+1}\,, \qquad (4.10a) \\
\triangle PQ_1 y_n &= \rho \triangle PQ_1 y_{n-1} + h\rho((PQ_1)'(t_{n-1}) + O(h))\triangle y_{n-1} + \\
&\quad (b^T\mathcal{A}^{-1} \otimes I_m) D_{PQ_1} \triangle \bar{Y} + h((b^T\mathcal{A}^{-1} \otimes I_m) D_{\gamma(PQ_1)'} + O(h))\triangle \bar{Y} + \\
&\quad hPP_1(t_n)\delta_{n,s+1}\,, \qquad (4.10b) \\
\triangle Q y_n &= \rho \triangle Q y_{n-1} + h\rho(Q'(t_{n-1}) + O(h))\triangle y_{n-1} + \\
&\quad (b^T\mathcal{A}^{-1} \otimes I_m) D_Q \bar{\Theta} + h((b^T\mathcal{A}^{-1} \otimes I_m) D_{\gamma Q'} + O(h))\triangle \bar{Y} + \\
&\quad hQ(t_n)\delta_{n,s+1}\,. \qquad (4.10c)
\end{aligned}
$$

Here, $\triangle PP_1 y_k = PP_1(t_k)(\tilde{y}_k - y_k)$, $\triangle PQ_1 y_k = PQ_1(t_k)(\tilde{y}_k - y_k)$, and $\triangle Q y_k = Q(t_k)(\tilde{y}_k - y_k)$, $k = n-1, n$.

Inserting (4.9a), (4.9b) into (4.10) and using the equation $\triangle \bar{Y} = D_{PP_1}\triangle \bar{Y} + D_{PQ_1}\triangle \bar{Y} + D_Q \bar{\Theta}$, we obtain the recursions:

$$
\|\triangle PP_1 y_n\| \leq \|\triangle PP_1 y_{n-1}\| + hC^{(n)}(\|\triangle y_{n-1}\| + \delta), \qquad (4.11a)
$$

$$
\|\triangle PQ_1 y_n\| \leq |\rho|\, \|\triangle PQ_1 y_{n-1}\| + hC^{(n)}\Big(\|\triangle y_{n-1}\| + \frac{\delta}{h}\Big), \qquad (4.11b)
$$

$$
\|\triangle Q y_n\| \leq |\rho|\, \|\triangle Q y_{n-1}\| + hC^{(n)}\Big(\|\triangle y_{n-1}\| + \frac{\delta}{h}\Big). \qquad (4.11c)
$$

Applying Lemma 4.1 to the recursions (4.11) and using the relation $\|\triangle y_n\| \leq \|\triangle PP_1 y_n\| + \|\triangle PQ_1 y_n\| + \|\triangle Q y_n\|$, we obtain

$$
\|\triangle y_n\| \leq C\delta. \qquad (4.12)
$$

Now the equation (4.9c) implies

$$
\|\triangle Z_j\| \leq \frac{C_1}{h}\Big(\|\triangle PQ_1 y_{n-1}\| + h\|\triangle y_{n-1}\| + \delta\Big).
$$

Inserting (4.12) into this relation, we obtain in the n-th integration step

$$
\|\triangle Z_j\| \leq \frac{C\delta}{h}, \qquad j = 1, 2, \ldots, s. \qquad \square
$$

To study the convergence we set $e_n^{(y)} = y_n - y(t_n)$, $E_j^{(y)} = Y_j - y(t_{nj})$ and $E_j^{(z)} = Z_j - z(t_{nj})$. Now, $e_n^{(y)}$ satifies the recursion

$$
e_n^{(y)} = \rho e_{n-1}^{(y)} + (b^T\mathcal{A}^{-1} \otimes I_m)(\bar{E}^{(y)} - \bar{\delta}^{(n)}) + \delta_{s+1}^{(n)}. \qquad (4.13)
$$

Under the assumptions of Theorem 4.1 the differences $E_j^{(y)}$ and $E_j^{(z)}$ satisfy

$$
D_{PP_1}\bar{E}^{(y)} = (D_{PP_1} + O(h))(\bar{e}_{n-1}^{(y)} + \bar{\delta}^{(n)}) + O(h\|\bar{\delta}_n\| + h\|\bar{\Theta}_n\|), \qquad (4.14a)
$$

$$
\begin{aligned}
D_{PQ_1}\bar{E}^{(y)} &= D_{PQ_1G_2^{-1}}\bar{\delta}_n\,, && (4.14b)\\
\bar{E}^{(z)} &= \Big[\frac{1}{h}(\mathcal{A}^{-1}\otimes I_m)D_{PQ_1P} - D_{QP_1G_2^{-1}BPP_1} + ((\mathcal{C}\odot\mathcal{A}^{-1})\otimes I_m)D_{(QQ_1)'}(D_{PP_1} - \\
&\quad I_{ms}) + O(h)\Big](\bar{e}^{(y)}_{n-1} + \bar{\delta}^{(n)}) + O\Big(\frac{1}{h}\|\bar{\delta}_n\| + \|\bar{\Theta}_n\|\Big), && (4.14c)
\end{aligned}
$$

where $\bar{E}^{(y)} := ((E^{(y)}_1)^T, (E^{(y)}_2)^T, \ldots, (E^{(y)}_s)^T)^T$, $\bar{E}^{(z)} := ((E^{(z)}_1)^T, (E^{(z)}_2)^T, \ldots, (E^{(z)}_s)^T)^T$, $\bar{\delta}^{(n)}$ and $\delta^{(n)}_{s+1}$ are as in (3.13a), $\bar{\delta}_n = (\delta^T_{n1}, \delta^T_{n2}, \ldots, \delta^T_{ns})$, $\bar{\Theta}_n = (\theta^T_{n1}, \theta^T_{n2}, \ldots, \theta^T_{ns})^T$, $\bar{e}^{(y)}_{n-1} := (e^{(y)}_{n-1} \otimes 1\!\!1_s)$, $e^{(y)}_{n-1} = y_{n-1} - y(t_{n-1})$. $(\delta_{nj}, \theta_{nj})$ are the numerical errors in the solution of the system (4.4) in the n-th integration step and $\delta^{(n)}_j$ is the j-th internal truncation error of the IRK method.

Theorem 4.3 *Let the assumptions of Theorem 4.1 be satisfied and suppose that for $q \geq 1$, $e^{(y)}_0 = O(h^{q+\beta})$ is the error in the initial condition and $\delta_{ij} = O(h^{q+\alpha})$ are the numerical errors in the solution of the system (4.4) in the i-th integration step. Then the global error e_n satisfies:*

- $\|e^{(y)}_n\| \leq Ch^q$ *if* $B(q)$, $C(q)$, $\alpha = \beta = 0$ *and* $|\rho| < 1$.
- $\|e^{(y)}_n\| \leq Ch^{q+1}$ *if* $B(q+1)$, $C(q)$, $\alpha = \beta = 1$ *and* $|\rho| < 1$.
- $\|E^{(z)}_j\| \leq Ch^q$ *if* $B(q)$, $C(q)$, $\alpha = \beta = 1$ *and* $|\rho| < 1$.
- $\|E^{(z)}_j\| \leq Ch^{q+1}$ *if* $B(q+1)$, $C(q)$, $b^T\mathcal{A}^{-1}c^{q+1} = 1$, $b^T\mathcal{A}^{-2}c^{q+1} = q+1$, $\alpha = \beta = 2$ *and* $|\rho| < 1$.
- $\|e^{(y)}_n\| \leq Ch^q$, $\|E^{(z)}_j\| \leq Ch^{q-1}$ *if* $\alpha = 1$, $\beta = 0$, $B(q)$, $C(q)$ *and* $|\rho| = 1$.
- $\|e^{(y)}_n\| \leq Ch^{q+1}$, $\|E^{(z)}_j\| \leq Ch^q$ *if* $\alpha = 2$, $\beta = 1$, $B(q+1)$, $C(q)$, $b^T\mathcal{A}^{-1}c^{q+1} = 1$ *and* $|\rho| = 1$, $j = 1, 2, \ldots, s$.

Proof As above we have the following recursions

$$
\begin{aligned}
\|PP_1e^{(y)}_n\| &\leq \|PP_1e^{(y)}_{n-1}\| + hC^{(n)}(\|e^{(y)}_{n-1} + D_1), && (4.15a)\\
\|PQ_1e^{(y)}_n\| &\leq |\rho|\,\|PQ_1e^{(y)}_{n-1}\| + hC^{(n)}\Big(\|e^{(y)}_{n-1} + \frac{D_2}{h}\Big), && (4.15b)\\
\|Q_1e^{(y)}_n\| &\leq |\rho|\,\|Q_1e^{(y)}_{n-1}\| + hC^{(n)}\Big(\|e^{(y)}_{n-1} + \frac{D_3}{h}\Big), && (4.15c)
\end{aligned}
$$

where

$$
D_1 = \begin{cases} O(h^q) & \text{if } \alpha = 0,\ B(q) \text{ and } C(q),\\ O(h^{q+1}) & \text{if } \alpha = 1,\ B(q+1) \text{ and } C(q), \end{cases}
$$

$$
D_2 = \begin{cases} O(h^q) & \text{if } \alpha = 0,\ B(q) \text{ and } C(q),\\ O(h^{q+1}) & \text{if } \alpha = 1,\ B(q) \text{ and } C(q),\\ O(h^{q+2}) & \text{if } \alpha = 2,\ b^T\mathcal{A}^{-1}c^{q+1} = 1,\ B(q) \text{ and } C(q), \end{cases}
$$

$$
D_3 = \begin{cases} O(h^q) & \text{if } \alpha = 0,\ B(q) \text{ and } C(q),\\ O(h^{q+1}) & \text{if } \alpha = 1,\ B(q) \text{ and } C(q),\\ O(h^{q+2}) & \text{if } \alpha = 2,\ b^T\mathcal{A}^{-1}c^{q+1} = 1,\ B(q) \text{ and } C(q). \end{cases}
$$

Applying Lemma 4.1 to the recursions (4.15) we obtain the statements of the theorem for the y-component. The statements for the z-component follow then from (4.14c) and the estimates for $e_n^{(y)}$. $\square$

Remarks

1. If $|\rho| < 1$, then Theorem 4.3 shows that the modified IRK method (4.2), (4.3) applied to the DAE (1.1) with variable nullspace is convergent with the same order if the IRK (1.2) is applied to the DAE (1.1) with constant nullspace.

2. For the case $|\rho| = 1$ we gain one power of h in the z-component by the modified IRK method (4.2), (4.3) compared to the statements of Theorem 3.4. This fact is not surprising because we have used the collocation equation (4.3a) instead of the recursion

$$z_n = \rho z_{n-1} + \sum_{j=1}^{s}\sum_{l=1}^{s} b_j \hat{a}_{jl} Z_l$$

   for the approximation of the z-component.

3. The conditions $|\rho| = 1$ and $b^T\mathcal{A}^{-1}c^{q+1} = 1$ are not fulfilled for any important method (for example the Gauss-Legendre formulas). So for these methods we must count with a loss of order of convergence. However, we can prove, in a manner analogous to the proof of Theorem 3.4, that $\|e_n^{(y)}\| \le Ch^{q+1}$ and $\|E_j^{(z)}\| \le Ch^q$ if $\alpha = 2$, $\beta = 1$, $B(q+1)$, $C(q)$ and $\rho = -1$.

5. Numerical experiments

Finally, we present the results of some numerical experiments which confirm the convergence statements predicted in this paper. The IRK method (2.1) and the modified IRK method (4.2), (4.3) have been implemented. In the experiments to determine the global error, we solved the problems with a sequence of fixed stepsize over the interval $[0, 1]$. Rates of convergence for each method were estimated by comparing the global errors at $t = 1$.

Example 5.1
The first test problem has constant nullspace. The problem is given by

$$\begin{aligned}
&e^{-t}(2x_2'(t) + x_3'(t)) + 2e^{-t}(\cos t x_1(t) + x_2(t) + (3\cos t - t - \frac{1}{2})x_3(t)) \\
&= 4 + e^{-t}\sin t(2e^{-t}\cos t - 1) + e^{-t}\cos t(6\cos t - 2t - 1),
\end{aligned}$$

$$\begin{aligned}
&\beta(t)(x_1'(t) + 3x_3'(t)) - \beta(t)(2e^{-t}x_2(t) - (6t + e^{-t})x_3(t)) + (t^2+1)x_3(t) \\
&= \beta(t)[(2e^{-t} + 6t)\cos t - (3 + e^{-t})\sin t - 2] + \cos t(t^2 + 1),
\end{aligned}$$

$$\beta(t)(x_1(t) + 3x_3(t)) = \beta(t)(e^{-t}\sin t + 3\cos t),$$

with $\beta(t) = \sin t + 2$.

For this problem, $Q = \begin{pmatrix} 0 & 6 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 0 \end{pmatrix}$ is a projection onto $N(t)$. With intial values given by $x_0 = (0, 1, 1)^T$ this problem has the solution

$$\begin{aligned} x_1 &= e^{-t} \sin t\,, \\ x_2 &= e^{t}\,, \\ x_3 &= \cos t\,. \end{aligned}$$

This problem has been solved using eight stepsizes $h \in H$ with

$$H := \{h \in \Re \;:\; \; h = 2^{-j} \text{ for } \; j = 3, 4, 5, \ldots, 10\}.$$

Table 5.1 gives the results of the experiments with methods of type (2.1).

Example 5.2
The matrices A and B are taken from Example 2.2 with the right-hand side

$$\begin{aligned} q_1(t) &= e^{-t}(\sin t + \eta t \cos t), \\ q_2(t) &= e^{-t}(\cos t - \sin t) - \eta t e^{-t}(\cos t + \sin t) + (1 + \eta) e^{-t} \cos t. \end{aligned}$$

This problem has the solution

$$\begin{aligned} x_1(t) &= e^{-t} \sin t, \\ x_2(t) &= e^{-t} \cos t. \end{aligned}$$

The problem has been solved for many values of η all between 0 and -100 by the modified IRK method. The method has been implemented for all stepsizes $h \in H$. Table 5.2 gives the results for $\eta = -1$. Note that e. g. the backward Euler formula in the non-modified form (2.1) leads to exponential instability.

Example 5.3
The following DAE is tractable with index 2 for all continuous differentiable funcions $\beta(t)$ and for $t \in [0, \infty)$.

$$\begin{aligned} e^{-t} x_2'(t) + \beta(t)(x_3'(t) + x_3(t)) + \beta'(t) x_3(t) &= (1 - e^{-t}) \sin t + \cos t, \\ 2x_1(t) + (\cos t + 2t + 2) x_3(t) &= 2e^t + e^{-t}(\cos t + 2t + 2), \\ 2x_1'(t) + (\cos t + 2t + 2) x_3'(t) &= 2e^t - e^{-t}(\cos t + 2t + 2). \end{aligned}$$

For this DAE,

$$Q = \begin{pmatrix} 0 & 0 & -\dfrac{\cos t}{2} - t - 1 \\ 0 & 0 & -\beta(t) e^t \\ 0 & 0 & 1 \end{pmatrix}$$

is a projection onto $N(t)$, and for $\beta(t) = e^t \sin t$ this problem has the solution

$$\begin{aligned} x_1(t) &= e^t, \\ x_2(t) &= \cos t, \\ x_3(t) &= e^{-t}. \end{aligned}$$

This problem has been solved by the modified IRK method which has been implemented for all the stepsize $h \in H$. Table 5.3 gives the results of the experiments.

Table 5.1

Method	ρ	K_d	K_I	$K_{a,1}$	EK		NK		$\underline{GF}$		$\overline{GF}$	
					P_x	Q_x	P_x	Q_x	Pe	Qe	Pe	Qe
1. midpoint rule	-1	2	1	1	2	0	2	0	6.4E-02	1.7E-00	2.5E-07	1.7E-00
2. backward Euler	0	1	1	∞	1	1	1	1	1.1E-01	2.6E-04	2.4E-04	7.3E-04
3. 2-Radau IIA	0	3	2	∞	3	2	3	2	3.4E-04	3.8E-03	2.6E-12	1.6E-08
4. 2-Lobatto IIIC	0	2	1	∞	2	1	2	1	3.0E-02	3.0E-01	1.2E-07	7.3E-04
5. 2-Radau IA	0	3	1	1	2	1	2	1	4.1E-02	5.3E-01	1.7E-07	1.5E-03
6. Crouzeix	ϑ	3	1	1	2	1	2	1	3.5E-02	2.4E-01	1.5E-07	7.0E-04
7. Alexander	0	2	1	∞	2	1	2	1	7.8E-04	2.2E-01	3.4E-09	5.2E-04
8. Gauss-Legendre	1	4	2	2	2	0	2	0	3.4E-03	2.5E-01	1.3E-08	2.5E-01
9. 3-Radau IIA	0	5	3	∞	5	3	5	3	4.1E-03	2.1E-02	1.7E-16	3.2E-10
10. 3-Lobatto IIIC	0	4	2	∞	4	2	4	2	5.2E-02	4.3E-01	2.5E-13	3.3E-05

explanations

1. midpoint rule or 1-stage Gauss-Legendre formula
2. backward Euler method
3. 2-stage Radau IIA
4. 2-stage Lobatto IIIC
5. 2-stage Radau IA
6. 2-stage A-stable SDIRK (Crouzeix)
7. 2-stage S-stable SDIRK formula (Alexander) with $\alpha = 1 - \frac{1}{2}\sqrt{2}$
8. 2-stage Gau-Legendre-Formel
9. 3-stage Radau IIA
10. 3-stage Lobatto IIIC

$-1 < \vartheta = -\frac{1+\sqrt{3}}{2+\sqrt{3}} < 1$

ρ	stability condition
K_d	nonstiff order
K_I	internal stage order
$K_{a,1}$	order of constant coefficient index 1 systems
EK	order predicted by the theorems 3.3, 3.4 and 4.3
NK	order of the observed global error
$\underline{GF}$	global error for the largest stepsize $h_{\max} = 2^{-3}$
$\overline{GF}$	global error for the smallest stepsize $h_{\min} = 2^{-10}$
Px	P-component
Qx	Q-component
Pe	P-component of the global error
Qe	Q-component of the global error

Table 5.2

Method	ρ	K_d	K_I	$K_{a,1}$	EK		NK		$\underline{GF}$		$\overline{GF}$	
					P_x	Q_x	P_x	Q_x	Pe	Qe	Pe	Qe
1. midpoint rule	-1	2	1	1	2	1	2	2**	6.5E-04	2.4E-03	2.3E-09	9.9E-09
2. backward Euler	0	1	1	∞	1	1	1	1	2.5E-21	1.9E-02	6.6E-24	3.8E-05
3. 2-Radau IIA	0	3	2	∞	3	2	0*	2	1.0E-20	2.5E-03	2.0E-20	1.3E-08
4. 2-Lobatto IIIC	0	2	1	∞	2	1	0*	1	6.8E-21	1.9E-02	6.8E-21	3.8E-05
5. 2-Radau IA	0	3	1	1	2	1	2	1	8.1E-04	1.3E-02	3.1E-09	2.6E-05
6. Crouzeix	ϑ	3	1	1	2	1	2	1	4.2E-04	1.1E-02	1.6E-09	2.2E-05
7. Alexander	0	2	1	∞	2	1	0*	1	6.8E-21	1.4E-02	1.4E-20	2.7E-05
8. Gauss-Legendre	1	4	2	2	2	1	2	1	1.4E-04	3.7E-03	5.2E-10	7.4E-06

$*$ In this case, the expected order of convergence does not yield. This is easy to explain by watching the global error, which lies within the computer precision limits.

$**$ The order of convergence is better then expected.

Table 5.3

Method	ρ	K_d	K_I	$K_{a,1}$	EK		NK		$\underline{GF}$		$\overline{GF}$	
					P_x	Q_x	P_x	Q_x	Pe	Qe	Pe	Qe
1. midpoint rule	-1	2	1	1	2	1	2	2*	5.9E-03	3.1E-02	2.3E-08	1.6E-07
2. backward Euler	0	1	1	∞	1	1	1	1	1.2E-01	1.9E+00	2.3E-04	4.0E-03
3. 2-Radau IIA	0	3	2	∞	3	2	3	2	2.6E-05	5.4E-02	1.9E-13	2.1E-07
4. 2-Lobatto IIIC	0	2	1	∞	2	1	2	1	1.9E-03	1.9E+00	7.1E-09	4.0E-03
5. 2-Radau IA	0	3	1	1	2	1	2	1	7.5E-03	1.1E+00	3.0E-08	2.6E-03
6. Crouzeix	ϑ	3	1	1	2	1	2	1	3.9E-03	7.7E-01	1.5E-08	2.3E-03
7. Alexander	0	2	1	∞	2	1	2	1	2.1E-04	1.3E+00	8.7E-10	2.8E-03
8. Gauss-Legendre	1	4	2	2	2	1	2	1	1.3E-03	3.5E-01	5.0E-09	7.6E-04

* The midpoint rule shows the same order in both components.

References

[1] **U. M. Ascher, L. R. Petzold** : (1991) *Projected implicit Runge-Kutta methods for differential-algebraic equations.*- SIAM J. Numer. Anal. 28, pp. 1097-1120.

[2] **K. E. Brenan, S. L. Campbell, L. R. Petzold** : (1989) *Numerical solution of initial value problems in differential-algebraic equations.*- North-Holland, New York.

[3] **J. C. Butcher** : (1987) *The numerical analysis of ordinary differential equations. Runge-Kutta and general linear methods.*- J. Wiley & Sons Ltd., Chichester.

[4] **E. Griepentrog, R. März** : (1986) *Differential-algebraic equations and their numerical treatment.*- Teubner-Texte zur Mathematik; Teubner, Leipzig.

[5] **E. Hairer, Ch. Lubich, M. Roche** : (1989) *The numerical solution of differential-algebraic systems by Runge-Kutta methods.*- Lecture Notes in Math.; 1409, Springer-Verlag.

[6] **M. Hanke** : (1990) *On the regularization of index 2 differential-algebraic equations.*- J. Math. Anal. Appl., 151(1), pp. 236-253.

[7] **M. Hanke** : (1992) *On the stability and convergence of one-step schemes for index 2 differential-algebraic equations.*- Preprint 92-11, Humboldt-Univ., Sektion Mathematik, Berlin.

[8] **E. Izquierdo**: (1992) *Numerische Approximation von Alrebro-Differentialgleichungen mit Index 2 mittels impliziter Runge-Kutta-Verfahren.*- Doctoral thesis, Humboldt-Univ., Berlin.

[9] **Kværnø, A.** : (1990) *Runge-Kutta Methods Applied to Fully Implicit Differential-Algebraic Equations of Index 1.*- Mathematics of Computation, vol. 54, N$\underline{\text{o}}$. 190, pp. 583-625.

[10] **R. März** : (1987) *Index-2 differential-algebraic equations.*- Preprint 135, Humboldt Univ., Sektion Mathematik, Berlin.

[11] **R. März** : (1987) *Higher-index differential-algebraic equations: Analysis and numerical treatment.*- Preprint 159, Humboldt Univ., Sektion Mathematik, Berlin.